CE LIBRARY

Vol. XVIII, No. 5 *PSYCHOLOGICAL REVIEW PUBLICATIONS* October, 1935

OCT 3 - 19

Journal of Experimental Psychology

EDITED BY

SAMUEL W. FERNBERGER, UNIVERSITY OF PENNSYLVANIA

WALTER S. HUNTER, CLARK UNIVERSITY (*Psychol. Index*)
HERBERT S. LANGFELD, PRINCETON UNIVERSITY (*Psychol. Rev.*)
JOSEPH PETERSON, GEORGE PEABODY COLLEGE (*Psychol. Monog.*)
JOHN A. McGEOCH, WESLEYAN UNIVERSITY (*Psychol. Bull.*)

CONTENTS

The Conscious Correlates of the Galvanic Skin Response: C. LANDIS AND W. A. HUNT 505

The Effect of the Time Interval upon the Time-Error at Different Intensive Levels: J. G. NEEDHAM 530

Simultaneous Lifting of Equally Heavy Weights by Both Right and Left Hands: N. SHEN 544

Fatigue, Oscillation, and Blocks: A. G. BILLS 562

The Influence of Relevance and Belonging: E. L. THORNDIKE AND I. LORGE 574

Retroactive Inhibition and Recognition Memory: F. McKINNEY 585

Backward Conditioned Eyelid Reactions: H. CASON 599

Influence of Syllabic Length and Rate of Auditory Presentation on Ability to Reproduce Disconnected Word Lists: S. W. CALHOON 612

A Comparison of Ability to Reproduce One-Syllable Words and Digits Auditorially Presented: S. W. CALHOON 621

An Experiment on Orientation in Stylus Maze Learning: H. N. PETERS AND L. McLEAN 633

A Comparison of Emphasis upon Right and upon Wrong Responses in Learning: W. S. HULIN AND D. KATZ 638

PUBLISHED BI-MONTHLY
FOR THE AMERICAN PSYCHOLOGICAL ASSOCIATION
BY THE PSYCHOLOGICAL REVIEW COMPANY
PRINCE AND LEMON STS., LANCASTER, PA.
AND PRINCETON, N. J.

Entered as second-class matter, July 1, 1920, at the post-office at Lancaster, Pa.

PUBLICATIONS

OF THE

AMERICAN PSYCHOLOGICAL ASSOCIATION

EDITED BY

S. W. FERNBERGER, University of Pennsylvania (*J. Exper. Psych.*)
WALTER S. HUNTER, Clark University (*Index and Abstracts*)
HENRY T. MOORE, Skidmore College (*J. Abn. and Soc. Psychol.*)
HERBERT S. LANGFELD, Princeton University (*Review*)
JOSEPH PETERSON, George Peabody College (*Monographs*)
JOHN A. McGEOCH, Wesleyan University (*Bulletin*)

HERBERT S. LANGFELD, Business Editor

PSYCHOLOGICAL REVIEW

containing original contributions only, appears bi-monthly, January, March, May, July, September, and November, the six numbers comprising a volume of about 540 pages.

PSYCHOLOGICAL BULLETIN

containing critical reviews of books and articles, psychological news and notes, university notices, and announcements, appears monthly (10 numbers), the annual volume comprising about 720 pages. Special issues of the Bulletin consist of general reviews of recent work in some department of psychology.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

containing original contributions of an experimental character, appears bi-monthly, February, April, June, August, October, and December, the six numbers comprising a volume of about 900 pages.

PSYCHOLOGICAL INDEX

is a compendious bibliography of books, monographs, and articles upon psychological and cognate topics that have appeared during the year. The Index is issued annually in June, and may be subscribed for in connection with the periodicals above, or purchased separately.

PSYCHOLOGICAL ABSTRACTS

appears monthly, the twelve numbers and an index supplement making a volume of about 700 pages. The journal is devoted to the publication of non-critical abstracts of the world's literature in psychology and closely related subjects.

PSYCHOLOGICAL MONOGRAPHS

consist of longer researches or treatises or collections of laboratory studies which it is important to publish promptly and as units. The price of single numbers varies according to their size. The Monographs appear at irregular intervals and are gathered into volumes of about 500 pages.

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY

appears quarterly, April, July, October, January, the four numbers comprising a volume of 448 pages. The journal contains original contributions in the field of abnormal and social psychology, reviews, notes and news.

ANNUAL SUBSCRIPTION RATES

Review: \$5.50 (Foreign, \$5.75). **Index:** \$4.00 per volume.
Journal: \$7.00 (Foreign, \$7.25). **Monographs:** \$6.00 per volume (Foreign, \$6.30).
Bulletin: \$6.00 (Foreign, \$6.25). **Abstracts:** \$6.00 (Foreign, \$6.25).
Abnormal and Social: \$5.00 (Foreign, \$5.25). Single copies \$1.50.
Current numbers: **Journal, \$1.25; Review, \$1.00; Abstracts, 75c; Bulletin, 60c.**

COMBINATION RATES

Review and Bulletin: \$10.00 (Foreign, \$10.50).
Review and J. Exp.: \$11.00 (Foreign, \$11.50).
Bulletin and J. Exp.: \$12.00 (Foreign, \$12.50).
Review, Bulletin, and J. Exp.: \$16.00 (Foreign, \$16.75).
Review, Bulletin, J. Exp., and Index: \$19.00 (Foreign, \$19.75).

Subscriptions, orders, and business communications should be sent to the

PSYCHOLOGICAL REVIEW COMPANY

PRINCETON, NEW JERSEY

Journal of
Experimental Psychology

VOL. XVIII, No. 5 OCTOBER, 1935

THE CONSCIOUS CORRELATES OF THE GALVANIC SKIN RESPONSE *

BY CARNEY LANDIS
Columbia University
AND
WILLIAM A. HUNT
Connecticut College

Two recent comprehensive reviews of the pertinent literature (**20**) (**17**) have shown the great diversity of opinion concerning the psychological correlates of the Galvanic Skin Response (GSR). The most numerous claims are those which maintain its correlation with emotion. The publications of Wechsler (**29**) and Ruckmick (**8**) are typical of this approach. Under Aveling (**2**) a British school has arisen that attempts to correlate the GSR with conation. Recently Cattell (**4**), having previously accepted the GSR as a measure of the conative tendency, included it in a battery of temperament tests. Bartlett (**3**), however, has denied the separability of feeling and conation in these introspections and designates the GSR as the accompaniment of the entire experience, or 'orectic tendency.' In this country, Abel (**1**) has labelled it as the accompaniment of 'certain reportable attitudes (Bewusstseinslagen) which mark sudden, decided and momentary checks in the course of · · · comprehension.' Other characteristic claims such as that of Grünbaum (**11**), who correlates it with

* This experiment was carried out in the Dept. of Psychology, N. Y. S. Psychiatric Institute and Hospital, during July and August, 1933. We desire to express our appreciation to the members of the staff of the Institute for their cooperation. We also are greatly indebted to Edna B. Hunt for her assistance in all parts of this work.

attention, and Weinberg (**30**), who related it to the raising of the level of consciousness, may be mentioned. As Landis (**18**) pointed out, this contradictory evidence would seem to indicate that the GSR may be a correlate of any variety of conscious reaction. However, each investigator has backed his conclusions with experimental evidence which he felt justified his position. Quite obviously, something must be wrong.

Several of the sources from which these differences of interpretation have arisen are apparent. The historical setting gives rise to some of the trends. No doubt the original interest of Vigouroux (**28**) in the electrical phenomena of the skin as a possible diagnostic instrument in clinical medicine directed attention toward the response as a possible correlate and measure of *something*. The early and somewhat fantastic attempts of Féré (**9**) to relate it to emotion, the subsequent claims of Veraguth (**27**) that it offered a method for investigation of emotional complexes, its definite inclusion in the work of the psychoanalytic school by Peterson and Jung (**22**) in 1907, all have played a large part in developing the persistent and tenacious idea, reflected in our modern textbooks, that this phenomenon is a definite correlate of emotion.

Difficulties of terminology also enter and add their perplexities. Such terms as emotion, conation, will, and attitude are so vague and generic as to lend themselves to all sorts of possible interpretations. The lack of specific definition and the consequent possibility of application to widely divergent phenomena certainly added their quota to the confusion. Certain problems involved in the technique of measuring these electrical changes have also increased the ambiguity of the results in the field. Many psychologists have rashly attempted the correlation of the GSR with various psychological factors with little or no idea of the physiological basis of the reaction and the technical hazards involved in its accurate measurement. Thus Forbes (**10**) has recently shown that the popular Hathaway (**12**) apparatus is an inadequate research instrument, although quite suitable for demonstration purposes.

Another of the chief reasons for these conflicting claims lies in the theoretical setting in which the experimentation

has been conducted. It is hard to evaluate the danger or extent of laboratory bias. This bias does not enter directly into the experimental procedure but manifests itself subtly in matter of interest, preliminary prejudicial influence of categories of report, interpretation, etc. It seems significant, that in the King's College laboratory the GSR should correlate with conation, in the Zurich psychoanalytic clinics with unconscious emotion, and at the Sorbonne with reportable emotion. The factors involved here no doubt play their part in the unfortunate narrow selection of stimuli that often occurs. If the GSR is a general response, it naturally would give an illusory correlation with conation or emotion if stimuli were used which favored one or the other of these types of reaction. To be adequate, the stimulus series should arouse a large range of psychological experiences. The selection of observers is also an important factor to be considered. A mere increase in number is not sufficient to insure a fair sampling when the group is drawn entirely within a certain sphere of systematic influence. We feel that a less uniform group of subjects would have improved many of the experiments reported in the literature.

Finally, some of the differences in the literature are no doubt due to logical difficulties. This has resulted in opposing inferences being drawn from the same data. Certain psychologists feel justified in calling the GSR a correlate of some psychological phenomenon if it can be shown to accompany that phenomenon even occasionally. Others feel that it must accompany that phenomenon *solely* and uniformly, if it is to be accepted as a correlate. Certainly the value of any measure as an indicator or practical method depends upon this last interpretation. Some insist upon its accompanying conscious, reportable emotion if it is to be accepted as an emotional index. Others avoid apparent discrepancies here by introducing the concept of unconscious emotion. The controversy over the definition of emotion is subtly reflected throughout the interpretation of the data obtained on the GSR. Certain of the conclusions of those who argue for the emotional specificity of the response are inferred upon physiological and neurological bases. Thus the mid-brain

and calamus scriptorius are said to be connected with general responses of an emotional nature. The GSR has been produced by direct electrical stimulation of these areas. Therefore, the GSR and emotion go together. Unfortunately, the GSR has also been elicited by electrical stimulation of the motor cortex, which serves 'voluntary' reactions entering into every psychological category. The fact that the response probably depends upon sweat gland activity, which activity is in turn controlled by sympathetic fibers, is also taken as a proof of its emotional nature, since sweating is a frequent concomitant of emotion. Here it must be pointed out that the more our knowledge of the autonomic nervous system develops the less certain we are of its place or importance in the emotional response. The attenuated basis of this inferential reasoning robs it of any great validity, although its persistent application has increased the ambiguity of our subject.

Purpose.—The purpose of the present experiment was the experimental investigation of the psychological correlates of the Galvanic Skin Response, with due regard to the critical factors mentioned above. It was felt that adequate control of these sources of error would make possible the obtaining of experimental data of a relatively definite conclusive nature. While the writers would prefer to be classified as psychological eclectics, it was felt that the primarily behavioristic background of the one and the primarily introspective background of the other insured a fair treatment of both approaches. An attempt was made to secure, not a large number of observers, but a widely divergent group representing different systematic approaches and different institutions. A group of psychopathic patients was also included for purposes of comparison. The stimuli were selected for their diversity. They covered a wide range of possible psychological experiences, and hence should offer a fair opportunity for any possible consciously reportable correlate to appear. No attempt was made to guide the observer's report into certain categories. In addition, the best available electrical techniques were used and definite care was exercised to keep the experimental procedure from bias.

Subjects. The normal subjects were Miss Bi., research physiologist; Mr. M., clerk; Mr. Gr., college student; Miss Ba. and Miss Go., graduate students in psychology; Dr. P. and Dr. S., clinical psychologists; Dr. Fo., research psychologist; Dr.'s L., J., and Ke., professors of psychology; Dr.'s H. and Ka., professors of psychiatry; and Dr. Fr., psychiatric interne. Systematically divided they consist of 2 naïve subjects, 3 students in advanced psychology, 1 introspectionist, 5 eclectics, and 3 psychoanalysts. Our observers had done graduate work leading to degrees in the psychological laboratories of 5 universities, while their affiliations at the time of the experiment add 2 more to the list. Thus the diversity of academic influences is fairly large. Of the psychopathic subjects, 4 were dementia praecox, and 4 were manic-depressive. (The diagnosis in each case was said to be clear and accurate by the medical staff.) The patients may be characterized briefly as follows: Manic-Depressive—Miss M., blue and depressed, suicidal thoughts, takes on burdens of the world; Miss H., blue and depressed, three previous suicidal attempts; Miss Hf., noisy and assaultive on admission, definite flight of ideas, crying and continual talking; Miss K., easily stimulated, flight of ideas, very loquacious, sexual fantasies resulting in false court charges. Dementia Praecox—Mr. K., extremely introverted, seldom talks and is very withdrawn, hears voices berating him, deteriorating rapidly; Mr. G., quiet and courteous but rather sluggish, neat and well-oriented, complains of lack of interest and energy, some apprehension about condition; Miss M., social and professional failure, delusions of persecution, sexual fantasies, legally committed case; Miss G., indifferent, apathetic, some deterioration, believes she is Mary, Queen of Scots, legally committed case.

Stimuli.—Three groups of stimuli were used. The first consisted of the following 30 words selected at random from the Jung (**16**) list of association words as adapted for use in English by Eder:

1. head	11. yellow	21. naughty
2. water	12. bury	22. brother
3. dead	13. habit	23. love
4. make	14. money	24. afraid
5. woman	15. finger	25. chair
6. cold	16. child	26. kiss
7. pond	17. sad	27. bride
8. sick	18. marry	28. bag
9. lamp	19. big	29. bed
10. jump	20. doctor	30. insult

As a second group we compiled the following check list of 20 association words.

1. sugar	11. tower
2. paper	12. abbreviate
3. bright	13. moon
4. window	14. corpse
5. amber	15. cow
6. murder	16. hypothesis
7. table	17. orange
8. rectangular	18. anger
9. dance	19. humiliation
10. blood	20. approximate

The third group of stimuli consisted of 30 assorted association words, questions,

tasks, colors, startle stimuli, jokes, etc. These 30 stimuli were chosen to represent 10 supposedly different psychological categories, and arranged in staggered order, thus.

1. 'blood'
2. 'inferiority'
3. 'castration'
4. "Where's that little Jewish boy that started to work this morning?" "Oh, he's left to go in business for himself."
5. "Name as many animals as you can think of." (After 15 seconds.) "Change to names of birds." (Subject halted after 15 seconds more.)
6. "Which do you prefer, violet or orange?"
7. "Would you say that fuel bears the same relation to combustion that stimulus does to nervous energy?"
8. Automobile horn sounded.
9. Olive green colored square shown for 10 seconds.
10. Puzzle picture shown for 15 seconds.
11. 'stink'
12. 'surpass'
13. 'lips'
14. "The blonde around the corner calls her sweetheart 'grapefruit,' because he is a little squirt."
15. "Name as many words ending in *d* as you can." (After 15 seconds.) "Change to words beginning with *r*." (Subject halted after 15 seconds more.)
16. "Add 13 to your present age and multiply by 2."
17. "Add 3 and 5 and 9 and 2 and 4 and 7 and divide by 13/16ths."
18. The experimenter shouts suddenly.
19. Yellow colored square shown for 10 seconds.
20. Puzzle picture shown for 15 seconds.
21. 'suicide'
22. 'defeat'
23. 'bull'
24. "The manager of a sideshow midget who had just died, called to console the widow during the wake. After a few remarks he asked to see the body of his small friend. 'Yes,' replied the widow, 'He's lying in the next room. But shut the door after you. The cat has dragged him in here three times today.' "
25. "Name as many adjectives beginning with *b* as you can." (After 15 seconds.) "Change to verbs beginning with *b*." (Subject halted after 15 seconds more.)
26. "Add the following numbers: 7, 11, 9, 3, 6, 8, 1, and 4."
27. "What is the capital of the United States? of France? of Montenegro?"
28. Experimenter slams stick on top of table.
29. Blue colored square shown for 10 seconds.
30. Puzzle picture shown for 15 seconds.

The colors used were standard Milton Bradley papers, 8″ by 8″ square. The puzzle pictures consisted of black forms against a white background, devised as a gestalt completion test by Street (**24**).

Stimuli 1, 11, and 21 were selected as words of general emotional stimulation value. Stimuli 2, 12, and 22 were selected as being Adlerian in nature. Stimuli 3, 13, and 23 were considered Freudian. Stimuli 4, 14, and 24 were supposedly humorous. Stimuli 5, 15, and 25 were of the sort used by Jones (**15**) to represent perseveration and its interruption. Stimuli 6, 16, and 26 resembled those used by Cattell (**5**) to arouse

conation. Stimuli 7, 17, and 27 were taken in part from Abel (1) and were designed to cause attitudes of check. Stimuli 8, 18, and 28 cause startle. It was hoped that stimuli 9, 19, and 29 might possibly arouse the affective response. Stimuli 10, 20, and 30 were designed to arouse gestalt 'completion.' The writers are fully aware of the difficulties involved here and realize that the categorical values of their stimuli cannot be expected to be perfect. That any three association words, for instance, could be expected to elicit an Adlerian interpretation in everyone of a group of subjects is too much to hope for. Nor could they hope that all the observers would find the jokes funny. Similarly, a few of the subjects found the proposition to be decided in stimulus 5 easy going and failed to experience any check or inhibition. It should be realized that the primary importance of these stimuli lies not in their objective classification but in the wealth of subjective experiences they produced as reported in the introspections. Their diverse nature makes it possible for them to arouse responses over the entire gamut of psychological experience. In the actual practice of the experiment they served very well. A brief series of similar stimuli was also devised and used in a preliminary training period with each subject during which no records were taken.

Apparatus.—A 2420c Leeds and Northrup reflecting galvanometer was used with a Wheatstone bridge. The arrangement of the bridge is that used by Darrow (**6**) in his Behavior Research Photopolygraph. The advantages of this arrangement as listed by Darrow (**6**) include high sensitivity, with a practically uniform, small current through all subjects regardless of the subject's resistance, and continuous quantitative records with uniform scale values for subjects having different resistance levels. A General Radio type 102-L Decade Resistance Box was used as the variable arm of the bridge, with one 1,000 Ω and three 10,000 Ω General Radio Resistance Coils, type 500, and one 100,000 Ω Sensitive Research Instrument Co. resistance coil used in the fixed arms. An Ayrton Universal Shunt across the galvanometer was used, although in the experiment itself the galvanometer was always used at full sensitivity. The deflections were followed by a manually operated pointer on a scale set 2 feet from the galvanometer mirror. This in turn was connected, by means of a 3⅓–1 reducing pulley, with a writing lever on a Weiss-Renshaw (**25**) polygraph. The records were measured through the use of a scale calibrated in units of 100 Ω. Also recording on the polygraph were a signal marker, a time marker beating seconds, and a tambour writing lever attached to a pneumograph. The galvanometer and polygraph, operated by an assistant, were in a room adjoining that in which the subject sat. A 40″ × 50″ window in the separating wall between the rooms contained a one way screen which enabled the subject to be observed from the recording room. The subject was seated in an adjustable chair with his back to the one way screen. An adjustable left arm of the chair served as a padded arm rest and held at one end two liquid, finger-cup electrodes of the zinc-zinc sulphate-sodium chloride type. Two straps held the subject's arm steady and discouraged unnecessary movements. Directly before the subject's chair, a large gray screen was mounted on the wall. The colors and pictures were exhibited through a 6″ × 6″ aperture in this. The experimenter sat at a table behind and at one side of the subject. Mounted on the table was the key controlling the stimulus marker and a very small signal light controlled by the assistant in the recording room. Reaction times were recorded to hundredths of a second by means of a stop watch. A recording thermometer and barograph gave continuous measures of temperature and barometric pressure in the experimental room.

Procedure.—The subject was seated in the chair and the first and third fingers of his left hand were washed with alcohol. These fingers were then dried and inserted into the saline of the electrodes and the straps were adjusted across the base of the back of

the hand and the lower forearm. He was told to sit as quietly as possible and not to move his fingers in the electrodes. The even, constant immersion of the fingers was facilitated by the fact that the second joint of each finger was crooked over the edge of the electrode. The hand was always in clear view, and any movement of either the hand or any other part of the body was observed and entered on the record by a special signal over the stimulus marker. The pneumograph was adjusted about the subject's chest in order that any deep breaths, which might give rise to apparent galvanic deflections, might be indicated on the record. After a wait of from 5 to 15 minutes, while the subject's resistance adjusted itself to a new level after the preliminary drop (the so-called 'curve of rest'), the experiment was begun. Three sittings of approximately an hour each were necessary. With one exception, these were spaced a day or more apart, depending upon the convenience of the observer.

The first hour served as a training period. During this hour the subject was given a series of stimuli equivalent to the various types used in the actual experiment and practice introspections were taken. No records were kept of this training period. At the close of this hour the check series of 20 stimuli were given and records taken. No introspections were called for, however. This gave a means of isolating the effect of introspection upon the galvanic response. For purposes of control, this check series was occasionally left out at the first sitting and run as a separate sitting between the second and third hours, or after the third. During the second hour the 30 words from the Jung list were presented and introspections taken after each stimulus word. These instructions were read, "Please sit as quietly as possible and keep your fingers steady in the electrodes. You will be given a series of association words. Respond to each with the first word that enters your mind. After each of your responses you will be asked to give a verbal description of what happened in your consciousness during the period of stimulation." No attempt was made to force any particular pattern of response upon these reports. If the subject asked for further instruction, he was told to give what seemed to him to be the most fitting description of his consciousness from time of stimulation to the giving of his response. During the third hour the series of mixed stimuli were presented. For this the instructions were changed as follows: "Please sit as quietly as possible and keep your fingers steady in the electrodes. You will be given a series of stimuli of various sorts. If single words are spoken, respond with the first word that enters your mind. If questions are asked, answer as best you can. If tasks are given you, please carry them out. After each of your responses you will be asked to give a verbal description of what happened in your consciousness during the period of stimulation." *In none of the sittings were new stimuli given or introspections called for until the subject's resistance had reached a stable level after the previous response.* This was signalled from the recording room by means of the previously mentioned small light. While this procedure added to the time involved it removed the possibility of overlapping curves and masking effects not adequately guarded against when a purely automatic time interval is used. The introspective reports were written down immediately as given, by the experimenter seated at the table beside the subject.

Several methods of scoring the galvanic response were considered and investigated experimentally (**14**). In view of our findings (Hunt and Hunt (**14**)) it was finally decided to use the absolute number of ohms of each deflection as a quantitative measure, and to supplement this with a statement of frequency of occurrence, in this case the number of times each stimulus produced a galvanic deflection expressed as a percentage of the total number of applications of the stimulus.

In presenting the experimental results, two complementary methods of treatment seem necessary. One, the objective approach, consists of a consideration of the galvanic responses with respect to the various types of stimuli presented. Thus, Table I presents the average galvanic scores of

TABLE I

30 ASSORTED STIMULI

Grouped According to the 10 Categories

Predetermined Categories	% Freq. of Occur.			*Av. No. of Ohms per Occur.		
	Nor.	D.P.	M.D.	Nor.	D.P.	M.D.
Emotional	81	66	92	489	333	525
				390	320	360
Adlerian	76	75	100	442	233	458
				390	190	320
Freudian	93	83	92	661	333	566
				540	290	470
Humor	79	66	92	492	250	422
				420	260	390
Perseveration	86	72	75	527	411	258
				420	260	220
Conation	74	75	53	442	377	103
				390	290	100
Check	76	66	58	370	400	225
				320	330	230
Startle	100	100	100	1018	522	1358
				540	370	660
Affectivity	61	66	42	145	192	58
				140	170	70
Completion	37	42	42	62 80	108 130	147 180

* The top cell in each segment contains the average number of ohms change with each GSR; the bottom cell represents the statistical average deviation (A.D.).

the 30 assorted stimuli grouped according to the 10 categories —emotional, Freudian, Adlerian, etc. The other, the subjective approach, treats the galvanic responses in relation to the conscious states reported by the observer in his introspections at the time of the reaction. Thus, Table II gives the average galvanic scores accompanying the various experiences reported, using the subjects' own categories of report. In Table III these reports are arbitrarily grouped into Affectivity (pleasantness, unpleasantness), Check (inhibition, uncertainty, confusion, unexpected), Conation (effort, attention, concentration), Emotion (emotion, relief, annoyance, afraid, scared, shocked, upset, embarrassment, apprehension, wonder, amusement, startle, surprise), Expectation (expectation, anticipation), Intellectuation, Tension, Miscellaneous (baffled, flabbergasted, bewilderment, blankness, puzzled, passive, etc., all terms which appeared once and which resist any but extremely arbitrary classification). It must be mentioned at this point that the check series of stimuli, given without introspections being called for, indicated that the deflections produced during the experiment were due to the stimulus situations and not to extraneous circumstances such as strain, embarrassment, effort, etc., introduced by the necessity of introspecting. The introduction of the task of introspection apparently did not influence the galvanic responses to the stimuli.

Table I shows the deflections produced by 30 stimuli of the 10 types. In this objective approach where the galvanic response is viewed as a direct correlate of the external stimulus situation (irrespective of any conscious contents present in the observer's experience) it must be emphasized that our objective classifications are arbitrary. Some of our emotional stimuli may not be genuinely emotional to certain subjects, etc. In general, however, the writers feel that the classification is justified. The first thing that is noticeable in our table is that the response was elicited by all forms of stimulation. All of our ten types of stimuli produce galvanic deflections. It would seem impossible to label the galvanic response as the exclusive correlate of any one type of stimulus. On the other

hand, one can easily see how the universality of the response might arouse illusory hopes in any investigation limited to one particular type of stimulation. While our results do not justify the conclusion that the galvanic skin response may follow any type of stimulus, this generalization is rendered extremely probable on the basis of the present findings—namely, that galvanic deflections may appear in response to stimuli of an emotional, Adlerian, Freudian, humorous, perseverational, conational, checking, startling, affective, or completion type. Thus Landis's (**18**) theoretical conclusion concerning the universality of the response is experimentally justified.

In comparing the relative effectiveness of the various types of stimuli for the production of the galvanic response our case is not so clear. The statistical average deviations as shown in Table I are quite large and render many of the differences between categories unreliable. A rank order grouping of the stimuli for the normal group gives us a descending order for frequency of occurrence from Startle with 100 percent down through Freudian, Perseverational, Emotional, Humorous, Adlerian, Check, Conational, Affective, to Completion with 37 percent. The rank order for the absolute number of ohms is practically the same. Another study (**21**) of these same data using other scoring techniques gave practically the same ranking of values. Inspection of this rank order would seem to indicate that on a relative basis, those stimuli primarily of an emotional nature usually produce galvanic responses, and produce responses of a greater amplitude. This is by no means to be interpreted as a statement by the authors that the galvanic response is a valuable index of emotion. It is pertinent at this point to ask whether or not it is valid to group 'startle' with emotion, or whether it should be considered independently. The question cannot be settled at present. The whole subject of the relation of startle to emotion is a vital one, however, and sooner or later must be faced experimentally as well as theoretically.

In discussing the two psychopathic groups we must remember the smallness of each group (4 patients in each)

and its highly selected nature. The rank orders for the abnormals agree fairly well with those for the normals. The correlation between normals and dementia praecox patients in rank order based on percentage frequency of occurrence of the response (Table I) is + .68 and on the absolute number of ohms is + .66; and between normals and manic-depressives, + .79 and + .80. The correlations between the two abnormal groups themselves are + .66 and + .39. This lower correlation between the two abnormal groups leads to further interesting relationships which the experiment brought out. If one examines Table I, one finds a tendency for the dementia praecox cases to be higher than the manic-depressives in their responses to stimuli of a conational and inhibitory nature (perseveration, conation, check), while the manic-depressives are higher in responses of an emotional sort (emotion, Adlerian, Freudian, humor, startle). There is a certain logic here that fits the conventional clinical description of these psychoses. However, we must remember that the small samplings, their selected nature, and the variability of response robs the results of anything but a suggestive value. Some explanation of the past confusion resulting from investigations of the clinical possibilities of the response in diagnosing the psychoses, which Landis (**19**) has previously pointed out, may be found here in the apparent tendency of the two groups to react differently to different types of stimuli. Two different investigators using two different types of stimuli might easily come to contradictory conclusions regarding the galvanic tendencies of dementia praecox and manic-depressive psychoses. Further investigation of the diagnostic possibilities of the response would do well to concern itself with relative differences in response to specific types of stimulation rather than confining itself to a general approach.

Let us now turn to the subjective approach as exemplified by Table II which gives the average scores for the deflections accompanying the conscious states reported by the Os in their introspections. It should be remembered that there was no attempt at guiding or forcing of reports into one or another

TABLE II

CONSCIOUS STATES REPORTED

	No. of Reports			% Freq. of Occur.			*Av. No. of Ohms per Occur.		
	Nor.	D.P.	M.D.	Nor.	D.P.	M.D.	Nor.	D.P.	M.D.
Effort	97	13	15	60	85	73	282	508	367
							280	420	490
Inhibition	64	2	14	81	100	100	403	950	643
							350	200	540
Amusement	41	2	4	85	100	100	605	750	1625
							360	50	610
Startle	40	2	9	92	100	100	900	850	1311
							560	650	720
Uncertainty	37	1	1	81	100	100	378	100	1000
							380	—	—
Tension	31	0	0	100	—	—	1248	—	—
							580	—	—
Intellectuation	24	4	2	71	100	100	317	375	600
							300	130	300
Confusion	25	0	1	92	—	100	804	—	200
							610	—	—
Unpleasantness	19	4	2	95	25	100	274	125	400
							140	380	300
Surprise	17	2	0	88	100	—	1018	500	—
							710	100	—
Pleasantness	14	1	6	64	100	17	164	200	17
							160	—	28
Relief	9	1	0	89	100	—	500	700	—
							340	—	—
Wonder	9	1	0	56	100	—	400	200	—
							420	—	—
Emotion	8	0	2	100	—	100	788	—	200
							290	—	0

TABLE II—(*Continued*)

	No. of Reports			% Freq. of Occur.			*Av. No. of Ohms per Occur.		
	Nor.	D.P.	M.D.	Nor.	D.P.	M.D.	Nor.	D.P.	M.D.
Annoyance	8	0	0	100	—	—	675	—	—
							310	—	—
Attention	8	0	0	50	—	—	125	—	—
							130	—	—
Expectation	7	1	0	71	100	—	443	1300	—
							360	—	—
Anticipation	6	0	0	83	—	—	600	—	—
							470	—	—
Shocked	6	0	0	100	—	—	966	—	—
							880	—	—
Apprehension	5	0	0	60	—	—	260	—	—
							170	—	—
Concentration	4	9	1	25	100	100	500	800	900
							750	380	—
Embarrassment	3	0	0	67	—	—	103	—	—
							120	—	—
Afraid	3	0	1	100	—	100	1100	—	210
							670	—	—
Scared	3	1	0	100	100	—	300	500	—
							0	—	—
Unexpected	3	0	0	33	—	—	100	—	—
							100	—	—
Upset	2	0	0	100	—	—	500	—	—
							400	—	—
Miscellaneous	29	0	0	76	—	—	496	—	—
							457	—	—

* Top cell in each segment contains the average number of ohms change with each GSR; the bottom cell the statistical average deviation (A.D.).

introspective category. The Os selected their own terms of report. Here again, as in Table I, the most striking thing is the universality of the galvanic response. It occurs as a correlate of all the conscious states listed (only those appearing at least twice in the reports are considered independently). Thus the universality of the deflections is not an artificial result of the arbitrary objective grouping used in the first table. While the listing in Table II covers only 26 varieties of conscious content (for the normal group) the range is sufficient to suggest that, just as the galvanic skin response might be said to be a possible correlate of any type of stimulus, so it may be a possible correlate of any sort of conscious experience. In short, it is not the exclusive correlate of anything, but rather would seem to be a phenomenon of great generality.

When we compare the various conscious states among the normals we find again, as in Table I, that those categories which would usually be spoken of as emotional states rank higher both in the frequency with which they are accompanied by deflections and in the amount of the accompanying deflections. The exaggerated size of the statistical average deviations renders many of the differences between individual categories unreliable, but an inspection of the scores of the separate categories reveals some interesting material. Whereas startle stimuli led the list in Table I, we find tension heading the scoring on Table II. Of course, the experience of tension enters into the startle situation, but it is interesting that the Os should report it as 'tension' rather than 'startle.' The reports of conscious experience of 'startle' are not as frequent as those of 'tension.' The outstanding score of 'tension' is one of the most interesting of the results of the subjective classification. There were 31 reports of tension. All of them were accompanied by deflections, and the average amount of the deflections was greater than that for any other category of reported conscious content. This is particularly interesting in view of Sears's (**23**) recent conclusion that ". . . perhaps tension or feeling-tension would best fit the cases of high psychogalvanic reactivity. . . ." It might be

possible to set up 'tension,' *not as an exclusive correlate*, but as that conscious experience best correlating with the galvanic skin response. Our results give some indication of this, and one can find a great deal of support in many previous experimental reports. In fact, in surveying the literature it is easy to read 'tension' into all the conscious correlates previously claimed for the galvanic skin response. It is always dangerous and sometimes hardly fair to reinterpret the published results of some other investigator, but one might say that Cattell's (**5**) voluminous data supported the case for tension better than that for conation. Thus in an informal investigation conducted by one of the present writers, there was a high correlation between the rank order of the conscious contents reported by Cattell arranged on the basis of the size of the accompanying galvanic deflections and their rank order as determined by the amount of 'tension' involved in them on the basis of judgments by 15 naïve judges. Unfortunately, there is a great deal of inference and interpretation involved in making out the case for 'tension' as a correlate of the galvanic skin response, and there has already been too much of this in the field. Nor is it easy to see the advantage of such a view. After all, the term 'tension' is so broad and covers such a multitude of possibilities ranging from tension in an isolated peripheral muscle to such intangibles as 'emotional' and 'psychic' tension that the benefits of introducing the term in an already confused field are very dubious. One is also tempted to theorize on the possible relationship existing between the experience of tension and our startle stimuli, both outstanding correlates of the GSR in our data. Once more the present benefits are dubious. At best these matters are indicative for further work.

There is also some interest in comparing pleasantness with unpleasantness. Unpleasantness is accompanied by deflections more often than pleasantness and the average size of the deflections is greater although the variability renders the difference unreliable. The tendency of unpleasantness to be more intense than pleasantness and to grade over into an emotional response has been reported by Hunt (**13**). These results bear on this question.

When we consider the subjective reports from our abnormal subjects we find ourselves definitely limited by the paucity of the introspective reports. Some of this is due to the limited size of the groups, and the difficulties of getting introspections from psychotic patients account for more of the trouble. 'Effort' is the only category of report appearing in both groups often enough for consideration, and here we find the dementia praecox cases scoring higher than the manic-depressives. This is in line with the usual clinical descriptions, but the large size of the statistical average deviations deprives it of validity.

In Table III we have the average scores of the subjective

TABLE III
CONSCIOUS STATES REPORTED
Arbitrarily Grouped

	No. of Reports			% Freq. of Occur.			*Av. No. of Ohms per Occur.		
	Nor.	D.P.	M.D.	Nor.	D.P.	M.D.	Nor.	D.P.	M.D.
Affectivity	33	5	8	82	40	38	219	163	250
							150	290	400
Check	129	3	16	82	100	100	521	225	314
							435	150	580
Conation	109	22	16	58	91	75	485	654	634
							387	400	695
Emotion	154	9	16	88	100	100	630	583	1618
							504	400	513
Expectation	13	1	0	77	100	—	600	1300	—
							470	—	—
Intellectuation	24	4	2	71	100	100	317	375	600
							300	130	300
Tension	31	0	0	100	—	—	1248	—	—
							580	—	—
Miscellaneous	29	0	0	76	—	—	496	—	—
							457	—	—

*Top cell in each segment contains the average number of ohms change with each GSR; bottom cell the statistical average deviation (A.D.).

reports from Table II, grouped arbitrarily under the categories of affectivity, check, conation, emotion, expectation, intellectuation, tension, and miscellaneous, some of which we used in Table I. This table supports the findings in Table II. Thus, for the normals we see tension leading the list, followed by emotion, both in frequency of occurrence and amount of deflection. The relation between affectivity and emotion is clearly shown in these grouped scores. While the affective consciousness is accompanied by deflections about as often as is the emotional one, the deflections are, in general, less. The fact that conation shows a generally low score certainly casts further doubt upon those who would claim it as the primary correlate of the galvanic response. Consideration of the abnormal responses in Table III is of course vitiated by the same paucity of response that marked the abnormals in Table II. In passing, we might note the tendency of the manic-depressives to rate considerably higher in the emotional category.

There are several incidental results of the investigation worth mentioning. The electrical resistance levels found among our observers should be put on record. The original values (before the rest period) of the skin resistance ranged from 24,620 Ω to 7,480 Ω with an average of 13,101 Ω among the normals; from 19,030 Ω to 9,470 Ω, average 13,843 Ω, for the dementia praecox cases; and from 22,160 Ω to 6,450 Ω, average 15,643 Ω, for the manic-depressives. After the waiting period during which the 'curve of rest' appeared, the resistance values ranged from 23,500 Ω to 5,800 Ω, average 12,708 Ω for the normals; from 12,600 Ω to 8,000 Ω, average 10,537 Ω, for the dementia praecox subjects; and from 30,800 Ω to 7,600 Ω, average 16,656 Ω for the manic-depressives.

While the introduction of the task of introspection did not seem to affect the deflections to the stimuli, the subsequent verbal reporting of these introspections was in itself apparently a frequent stimulus for the galvanic skin response. Out of 803 introspections for the normals, 487 were accompanied by deflections. The majority of these were slight, but 27 of them ranged over 1,000 Ω. For the dementia praecox group there

were 225 introspections with 157 deflections, 7 of them over 1,000 Ω. For the manic-depressives there were 222 introspections with 152 deflections, 14 of them over 1,000 Ω. The types of conscious content contained in the introspections accompanying these deflections varied considerably but in general conformed to the distributions reported in Tables II and III. Thus, among the other correlates of the galvanic skin response it is necessary to mention the verbal reporting of introspections.

The writers (**21**) have already reported on the relation between the size of the galvanic deflections and the length of the reaction time on the word-association list, finding a slight tendency for the longer reaction times to be accompanied by the larger galvanic deflections. The relation between these measures and that third complex-indicator, the appearance of individual word responses on the association list, is of interest. The responses on the list of 30 association words were scored according to the Kent-Rosanoff (**26**) technique and standards. Since a word simply *is* or *is not* scored as an individual response there is no measure of degree of individuality or abnormality. Hence, the writers simply inspected the individual word responses to see whether or not they were accompanied by significant galvanic deflections or reaction times. As the criterion of significance the writers arbitrarily selected the standard deviation. All the galvanic deflections which were one sigma or more above the mean of the series were selected as significant. All the reaction times lying one sigma either above or below the mean of their series were selected as significant, since an exceedingly short reaction time is supposed to be as significant as an over-long one in the association test. Upon inspection we found 116 individual responses given by the normals. Of these, 29 were accompanied by galvanic deflections above the standard deviation, 38 were accompanied by reaction times above the standard deviation, and 5 by reaction times below the standard deviation. Among the dementia praecox cases there were 24 individual responses, 2 accompanied by significant deflections, 7 by long reaction times, and 1 by a short reaction time.

The manic-depressives gave 33 individual responses, 6 accompanied by large deflections, 11 by long reaction times, and 1 by a short reaction time. The difficulty of defining a complex-indicator and the roughness of our measures render any conclusions extremely tentative, but here again, as in the case of galvanic magnitude and reaction time, there would seem to be a very slight tendency for these two measures to agree with the third, namely the occurrence of individual word responses. The very brief reaction time seems to be least in agreement with our other indicators, particularly since none of them occur with large galvanic deflections, whereas 24 of the long reaction times do.

The bi-serial r correlation formula (Dunlap and Kurtz (7)) is designed to give a measure of relationship when one of the two variables presents a dichotomic grouping. In these data then we have the dichotomy 'significant-non-significant' response related to both the size of the GSR in ohms of change, and reaction-time in seconds. The r between 'significance' and GSR is $+ 0.22 \pm .03$; and with reaction-time is $+ 0.49 \pm .02$. These figures show the same thing which we have stated above, namely, that there is a tendency for these measures to occur together.

So far we have been speaking of the *galvanic skin response* as the response to our various types of stimuli and correlate of the different reported conscious contents. It is a pertinent question to consider whether this entire viewpoint might better be abandoned and to substitute the phrase *participation of the sympathetic division of the autonomic nervous system* for the term GSR. Since the galvanic skin response is generally accepted today as being a phenomenon directly related to sympathetic activity it would seem that there are advantages in referring directly to this, rather than viewing the deflections as isolated phenomena, important in their own right. This tendency to view the galvanic response as a more or less independent phenomenon has no doubt added to the confusion in the field. There is a biological definiteness about the sympathetic division of the autonomic nervous system and enough is known about it to prevent at least some of the

wilder theoretical outrages perpetrated against the galvanic skin response *per se*.

Let us then reinterpret the more important of the above results in terms of the action of the sympathetic division of the autonomic nervous system. We find, then, that the outstanding fact is the general participation of the sympathetic system in the behavioral picture. It would seem to be involved in our responses to all types of stimuli, and to be a possible neurological correlate of all our types of conscious content. The sympathetic nervous system may be said to have a tendency to be involved in any type of behavior. It is specific to none of the categories selected in this experiment. Indeed, this is what one logically might suspect. The autonomic nervous system is too intimately connected with the basic physiological processes, too closely related to the central nervous system, and there is too close an interrelationship between its various segments to allow for the functional segregation of the entire system or any one of its divisions for exclusive use in any one type of reaction. Thus to assume that the thoracico-lumbar system might be exclusively utilized by unpleasant emotion, as some theoretical writers have done, is flying in the face of all common sense. The only way such a view could be justified is by the dialectical expedient of defining emotion *as* activity of this division of the autonomic system.

We must point out, however, that while this sympathetic division participation is general and not the exclusive correlate of any specific stimulus or experience, it may be differentiated on a quantitative basis. Here we find that those stimuli which we presume might give rise to emotion, either subjective or objective, are usually among the most effective in setting up changes in the amount of sympathetic activity. The best type of stimulus for this is anything startling. Here again these conclusions are in line with those from the other work in the field. The conscious content reported most frequently when the largest changes of sympathetic activity took place is that called by the observers 'tension.' However, until this term achieves a more exact reference, its significance is

dubious. Finally, the fairly high score of conscious states which may be classified as affection, in frequency of occurrence and their relatively low score in actual ohms deflection, suggests that the affective states resemble the emotional ones in that they usually involve the sympathetic nervous system, but might be differentiated from them on the basis of the minor magnitude of this participation. Unpleasantness seems to involve more sympathetic activity than pleasantness.

Discussion and Conclusions

We have attempted a critical experimental evaluation of the possible adequate stimuli and accompanying conscious correlates of the galvanic skin response. We have set up the conditions of the experiment so that we might benefit from the experience of previous investigators. We have tried to make our methods adequate and clear-cut. We made use of carefully selected groups of observers who were chosen to fit into particular categories of training, systematic bias, naïveté, psychoses, etc. We chose our stimuli to cover both the original claims of Jung (**16**) and his associates, from whose work so much has been generalized in the literature, and the present day psychological classificatory categories. We have made use of the best standardized method for the electrical measurement and recording of this phenomenon (Darrow). We have evaluated four possible ways of scoring the GSR data. We have analyzed the effect of systematic training or bias.* We have been scrupulously careful in the classifying of our reports in the obvious terms with which they were given. It is to be remembered that we did not ask our observers to use, nor did we train them to, any standard variety of report. Having thus carefully attempted to obtain a clear-cut set of observations, what are the results?

We find clear and unmistakable evidence that the GSR may occur after any one of a very wide variety of stimuli or situations. It may accompany *any* word in the 'diagnostic association word list,' whether that word is accompanied by

* Cf., Hunt, W. A., and Landis, C., Systematic differences in introspection, *Amer. J. Psychol.*, 1934, **46**, 659–660.

other 'complex indicators' or not. There is a tendency for the GSR to go with other 'complex indicators,' *but only a tendency*. The GSR may accompany any variety of introspective report given by any observer of any sort of systematic training or lack of training. It has a tendency to occur more frequently when either tension or startle is reported as the conscious correlate. It may go with the giving of any sort of introspection. And lastly, it sometimes occurs 'spontaneously,' that is, with no known stimulus or consciously reportable concomitant.

After all, is not this just what should be expected? The GSR is a phenomenon reflexly controlled by the sympathetic division of the autonomic nervous system, just as the knee jerk is controlled through the lumbar section of the spinal cord. Everyone knows that either 'startle' or 'tension' (in the sense of readiness to react) increases the extent of reflexly controlled action, so it is with the GSR. In that variety of psychological experience called emotion or affection, we very frequently find elements of tension or startle. It is true that emotion which is sustained more than a very few minutes usually carries with it no element of tension or startle; hence one would not expect the GSR to appear frequently in sustained emotion. The entire question of the relationship between startle, tension and emotion in general, is one which, psychologically, must be investigated further. The statement that startle is a simple or initiatory form of emotion is not an adequate explanation.

If we use the term 'conation' as Aveling (**2**) has done, namely, to describe any variety of mental activity involving volition or willing, then it is apparent that the degree to which conational activity incorporates the element of tension (readiness to react) will be reflected in an increased frequency of occurrence of the GSR.

Bartlett (**3**) has spoken of the possibility that the GSR might be associated with 'complex orectic tendencies,' at least to the degree that reactivity in general enters into desire or appetition. Unfortunately, neither our experiments nor those of our predecessors have demonstrated an exclusive

concomitant stimulus or conscious response for the GSR. It goes with startle, emotion, desire, and so forth in a fair percentage of instances, but not invariably so.

It seems clear, then, that this response is *not* associated to any marked degree with any one variety of stimulus or conscious state. It seems to be more nearly related to startle or to tension than to anything else. However, we do not feel that there is any evidence which would justify our saying more than that the GSR is a fairly adequate indicator of 'change of direction of mental activity,' but in no sense is it an adequate or direct measure of these changes.

Summary

In summary, a group of observers differing widely in psychological experience and systematic bias and including some psychopathic cases were submitted to a wide range of stimulus situations while their galvanic skin responses were being measured. They were asked to introspect on the conscious experiences accompanying these stimuli. The results show that the galvanic skin response accompanies all the types of stimuli used and all the types of conscious content reported. Relative quantitative differences do exist, however, and we find emotion consistently high, both in the number of deflections occurring with it and in the size of the deflections. It is suggested that these results be interpreted in terms of the participation of the sympathetic division of the autonomic nervous system.

(Manuscript received July 26, 1934)

References

1. Abel, T. M., Attitudes and the galvanic skin reflex, J. Exp. Psychol., 1930, **13**, 47–60.
2. Aveling, F., The psychogalvanic phenomenon, Proc. IX Int. Cong. Psychol., 1929, **9**, 63–65.
3. Bartlett, R. J., Does the psychogalvanic phenomenon indicate emotion? *Brit. J. Psychol.*, 1927–28, **18**, 30–50.
4. Cattell, R. B., Temperament tests. II. Tests, *Brit. J. Psychol.*, 1933, **24**, 20–49.
5. Cattell, R. B., Experiments on the psychical correlate of the psychogalvanic reflex, *Brit. J. Psychol.*, 1928–29, **19**, 357–386.

6. Darrow, C. W., Studies in the dynamics of behavior. Chicago, Univer. of Chicago Press, 1932, 258–261.
7. Dunlap, J. W., and Kurtz, A. K., Handbook of statistical nomographs, tables, and formulas. Yonkers, World Book Co., 1932, 46–47.
8. Dysinger, W. S., and Ruckmick, C. A., The emotional responses of children to the motion picture situation, New York, Macmillan, 1933, pp. 122.
9. Féré, C., The pathology of emotions. (Translated by Park, R.) London, University Press, 1899, 164–166.
10. Forbes, T. W., The Hathaway galvanic apparatus as an instrument of research. *Amer. J. Psychol.*, 1933, **45**, 502–511.
11. Grünbaum, A. A., Aufmerksamkeit, Emotivität und galvanisches Phänomen bei morbus Basedowi, *Psychiat. en neurol. Bladen*, 1922, **26**, 359–371.
12. Hathaway, S. R., A comparative study of psychogalvanic and association time measures—a new psychogalvanic apparatus, *J. Appl. Psychol.*, 1929, **13**, 632–646.
13. Hunt, W. A., The pressure correlate of emotion, *Amer. J. Psychol.*, 1931, **43**, 600–605.
14. Hunt, W. A. and Hunt, E. B., A comparison of five methods of scoring the galvanic skin response, J. Exp. Psychol., 1935, **18**, 383–387.
15. Jones, L. W., An investigation into the significance of perseveration, *J. Ment. Sci.*, 1928, **74**, 653–659.
16. Jung, C. G., Studies in word-association (translated by Eder, M.D.). London, Heinemann, 1918, 446–530.
17. Landis, C., Electrical phenomena of the skin (Galvanic skin response), *Psychol. Bull.*, 1932, **29**, 693–752.
18. Landis, C., Psychology and the psychogalvanic reflex, *Psychol. Rev.*, 1930, **37**, 381–398.
19. Landis, C., Psychiatry and the "psychogalvanic reflex," *Psychiat. Quart.*, 1932, **6**, 262–272.
20. Landis, C., and DeWick, H. N., The electrical phenomena of the skin (Psychogalvanic reflex), *Psychol. Bull.*, 1929, **26**, 64–119.
21. Landis, C., and Hunt, W. A., Word-association reaction-time and the magnitude of the galvanic skin response, *Amer. J. Psychol.*, 1935, **47**, 143–145.
22. Peterson, F., and Jung, C. G., Psycho-physical investigations with the galvanometer and pneumograph in normal and insane individuals, *Brain*, 1907, **30**, 153–218.
23. Sears, R., Psychogalvanic responses in arithmetical work: Effects of experimental changes in addition, *Arch. of Psychol.*, No. 155, 1933, 56–57.
24. Street, R. F., A gestalt completion test, Bureau of Publications, No. 481, Teachers College, Columbia University, 1931, pp. 34.
25. Renshaw, S., and Weiss, A. P., Apparatus for measuring change in bodily posture, *Amer. J. Psychol.*, 1926, **37**, 261–267.
26. Rosanoff, A. J., Manual of psychiatry, New York, Wiley, 1920, 547–620.
27. Veraguth, O., Das psychogalvanische Reflex-phänomen, Berlin, Karger, 1909.
28. Vigouroux, R., Sur le rôle de la résistance électrique des tissus dans l'électrodiagnostic, *Gazette méd. de Paris*, 1879 (6th Ser.), **1**, 657–658.
29. Wechsler, D., The measurement of emotional reactions, *Arch. of Psychol.*, No. 76, 1925, pp. 181.
30. Weinberg, A. A., Psyche und unwillkürliches Nervensystem, *Zsch. f. d. ges. Neurol. u. Psychiat.*, 1923, **85**, 543–565; **86**, 375–390; 1924, **93**, 421–445.

THE EFFECT OF THE TIME INTERVAL UPON THE TIME-ERROR AT DIFFERENT INTENSIVE LEVELS[1]

BY J. GARTON NEEDHAM[2]

The time-error is a measure of the relative over-estimation (preponderance of judgments 'greater') or under-estimation (preponderance of judgments 'less') of a given series of comparison stimuli presented for differential judgment. A significant characteristic of the time-error is its *p*-function, a term which refers to the variation of the time-error as there is a change in the time interval between the two members of a pair of stimuli which are to be compared.

The *p*-function does not represent the only dimension in which the time-error may vary. A number of experimental studies have indicated the nature of the displacements undergone by the time-error as the result of 'pre-disposing' effects of certain other stimulus conditions extraneous to the stimuli judged; the term 'contrast' is useful as an approximate description of these effects.

Woodrow's experiment,[3] which was carried out under the conditions of the comparison judgment, served to indicate the general nature of these 'contrast' effects. Within a series of stimulus presentations Woodrow included pairs (standard and comparison stimuli) of haphazardly varying intensive level. For example, the comparison of a weight of 113.3 grams with a standard of 110 grams would be followed by the comparison of 194 grams with a standard of 200 grams. The construction of the comparison series was exactly the same with each of the varying standards, but the effect of the experimental conditions was a regular variation in the 'time-order error' for the different intensive levels. The relative over-

[1] From the Princeton Psychological Laboratory.

[2] National Research Council Fellow.

[3] H. Woodrow, Weight discrimination with a varying standard, *Amer. Jour. Psychol.*, 1933, **45**, 391–416.

estimation (negative time-error) of the heavier stimulus-pairs became an under-estimation (positive time-error) in the case of the lighter pairs of stimuli.

Pratt [4] and Needham [5] have attempted to differentiate the extent to which such effects as those noted by Woodrow are inherently characteristic of the stimulus level and to what extent they result from the 'pre-disposing' effects of the experimental conditions. The second alternative is to be favored, especially if the first case is made a special instance of the second by the very reasonable assumption of a 'general level of reference' established from the range of past experience.

The studies of Pratt and Needham were carried out by the Method of Single Stimuli (absolute judgment), and consequently they provided no direct information as to the effect of the time interval upon the judgment accorded to the second of a pair of comparison stimuli. The work of Köhler [6] upon this problem is the point of departure for modern discussion; Woodrow revised and adapted Köhler's theories to accord with his data. It is interesting to note that the time-error for the comparison judgment follows an almost identical course under the two sets of experimental conditions, the variation of the time interval (Köhler) and the interspersion of different intensive levels (Woodrow). This approximation is indicated in Fig. 1, which shows curves constructed from the data of Köhler [7] and of Woodrow.[8]

Köhler's theoretical discussion is concerned with the dynamic changes undergone by the 'trace' of the first (standard) stimulus during the interval which intervenes before the appearance of the second stimulus, and concludes that these changes are in the nature of a 'sinking.' Woodrow portrays this trace as a 'set,' an expectation of a stimulus of a certain intensity. The *p*-function findings (Köhler's results, for

[4] C. C. Pratt, Time-errors in the method of single stimuli, Jour. Exper. Psychol., 1933, **16**, 798–814.

[5] J. G. Needham, Contrast effects in judgments of auditory intensities, Jour. Exper. Psychol., 1935, **18**, in press.

[6] W. Köhler, Zur Theorie des Sukzessivvergleichs und des Zeitfehler, *Psychol. Forsch.*, **4**, 1923, 115–175.

[7] *Loc. cit.*, p. 151.

[8] *Loc. cit.*, p. 400.

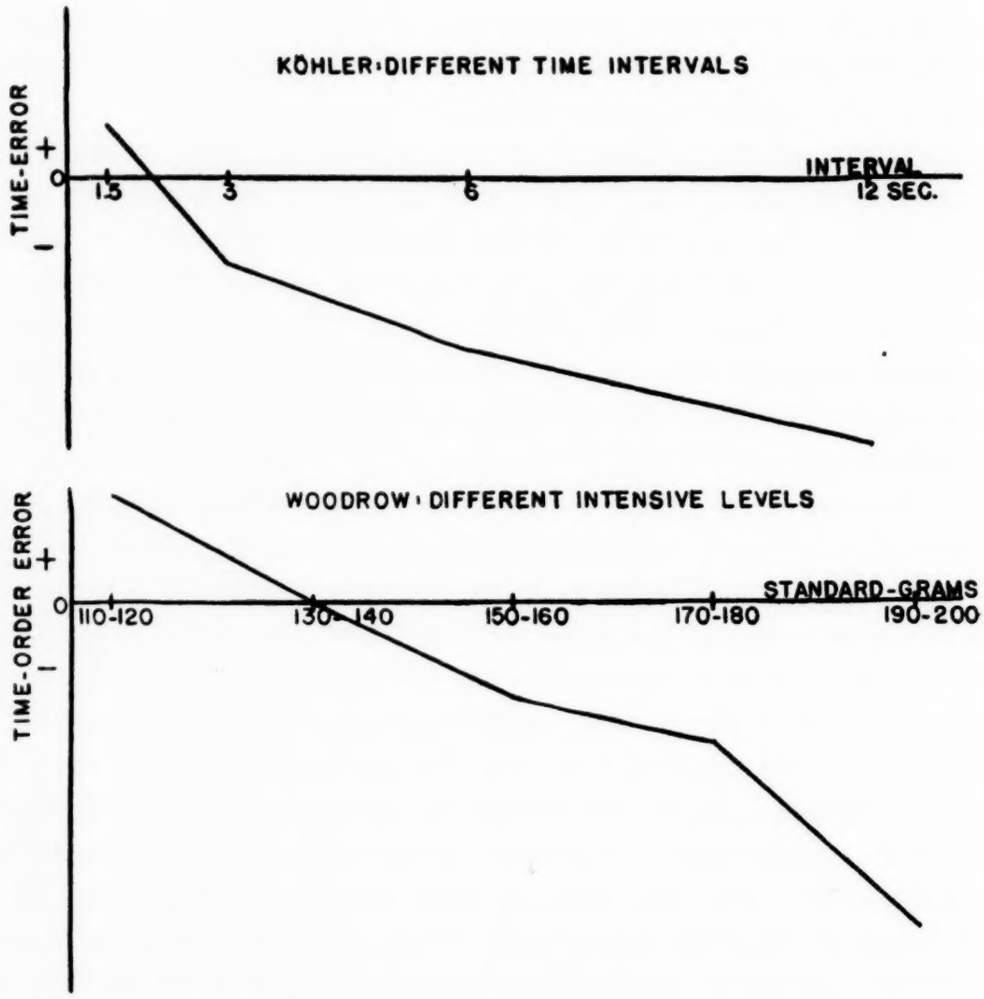


FIG. 1. Curves drawn from the data of Köhler and Woodrow (see text for references.)

example) are referred to a general 'sinking' of this 'trace-set,' while Woodrow's own results are explained by the postulation of an assimilation of the 'trace-set' of any given variable stimulus to a 'weighted general level' of all previous stimuli. This assimilation would be in opposite directions according as the particular standard was heavier or lighter than the average or general level involved, and from this the appearance of over- or under-estimation at the opposite extremes of the scale would follow logically.

A question arises as to the course of the process, whether its nature be one of assimilation, of contrast, or of anything else, when there is superimposed upon Woodrow's experimental conditions a variation of the time-interval. The ex-

periments which are reported here were concerned with this problem.

I

The first experiment involved comparison judgments of pairs of stimuli of varying intensive level with the inclusion of a variation of the time interval separating the members of the stimulus-pairs. The stimuli were auditory intensities. An electrically generated alternating current was led through an attenuator into an earphone at the right ear of the observer, producing a pure tone of 1100 d.v./sec. Series of standard-comparison stimuli were established at four levels of intensity. Each series comprised a standard stimulus, these being respectively 30, 40, 50, and 60 db above an approximated RL, and five comparison stimuli; in each case the comparison stimuli were ±2 db, ±1 db, and 0 db different from the standard stimulus. This selection was intended to provide, as far as possible, separations of both levels and comparison stimuli which were phenomenally equal,[9] and levels which ranged from weak to moderately loud.

Time controls were provided by a modified Lewin-Schumann *Zeitsinnapparat.* All stimuli were of 1 second duration. The intervals between the comparison stimuli were 1, 3, and 6 seconds. Each experimental hour was divided into three periods, and a different interval was used in each of these periods. The order in which the intervals were employed varied from one observer to another and from one experimental session to the next.

The three observers were Dr. G. P. Horton (*Ho*), a member of the laboratory staff, and Messrs. F. V. Taylor (*Ta*) and S. W. Cummings (*Cu*), graduate students in psychology. In each period of the five experimental sessions an observer made five judgments upon each of the twenty stimulus-comparison pairs (five comparison stimuli with each of four standards), which were presented in an order which was haphazard with respect to both intensive level and position within the comparison series. This procedure provided a total of 125 reports upon each comparison series at each of the three time intervals.

Three categories of judgment were permitted; the observers were requested to report the second member of each comparison pair as 'Louder,' 'Equal,' or 'Weaker' than the first. Table 1 presents the respective frequencies of judgment at each of the four intensive levels and at each time interval.

The D percent is a satisfactory measure of the time-error. Derived from the formula D percent $= (W - L)/N$, it indicates directly the relative preponderance of the one or the other class of judgments. Table 2 presents the values of the D percent calculated from the data in Table 1. The graphic presentation of Fig. 2 is based upon the values of the D percent derived from the total judgments of the three observers.

[9] The decibel is considered most adequate as a sensation unit for frequencies of the order of 1000 d.v./sec. Cf. H. Fletcher, *Speech and Hearing*, 1929, p. 69; H. Fletcher and W. A. Munson, *Jour. Acous. Soc. Amer.*, 1933, **5**, pp. 84f.

TABLE 1

TOTAL FREQUENCIES OF JUDGMENTS UPON COMPARISON STIMULI AT EACH OF FOUR INTENSIVE LEVELS AND THREE TIME INTERVALS

Observer	Interval	Intensive Level											
		60 db			50 db			40 db			30 db		
		L	E	W	L	E	W	L	E	W	L	E	W
Ta.....	1 sec	47	18	60	41	21	63	37	54	34	40	51	34
	3 sec	58	17	50	54	18	53	49	28	48	36	35	54
	6 sec	68	6	51	64	9	52	43	16	66	20	26	79
Cu.....	1 sec	35	43	47	41	43	41	29	63	33	19	66	40
	3 sec	49	55	21	50	42	33	33	48	44	22	63	40
	6 sec	73	34	19	65	32	28	28	55	42	13	33	79
Ho.....	1 sec	46	41	38	43	40	42	37	60	28	30	53	42
	3 sec	56	41	28	34	50	41	34	52	39	9	75	41
	6 sec	60	40	25	41	47	37	28	51	46	21	40	64
Totals (3 Os)...	1 sec	128	102	145	125	104	146	103	177	95	89	170	116
	3 sec	163	113	99	138	110	127	116	128	131	67	173	135
	6 sec	201	79	95	170	88	117	99	122	154	54	99	222

An inspection of Table 2 will show that there was a high amount of consistency among the individual observers, and that the use of the totals preserves the significant features of the data.

The results of this experiment, as portrayed in Fig. 2,

TABLE 2

VALUES OF THE D PERCENT DERIVED FROM TABLE 1

Observer	Interval	Intensive level			
		60 db	50 db	40 db	30 db
Ta.....	1 sec	+10.4	+17.6	− 2.4	− 4.8
	3 sec	− 6.4	− 0.8	− 0.8	+14.4
	6 sec	−13.6	− 9.6	+18.4	+47.2
Cu.....	1 sec	+ 9.6	0	+ 3.2	+16.8
	3 sec	−22.4	−13.6	+ 8.8	+14.4
	6 sec	−43.2	−29.6	+11.2	+52.8
Ho.....	1 sec	− 6.4	− 0.8	−7.2	+ 9.6
	3 sec	−22.4	+ 5.6	+ 4.0	+25.6
	6 sec	−28.0	− 3.2	+14.4	+34.4
Totals.....	1 sec	+ 4.5	+ 5.6	− 2.1	+ 7.2
	3 sec	−17.1	− 2.9	+ 4.0	+18.1
	6 sec	−28.3	−14.1	+14.7	+44.8

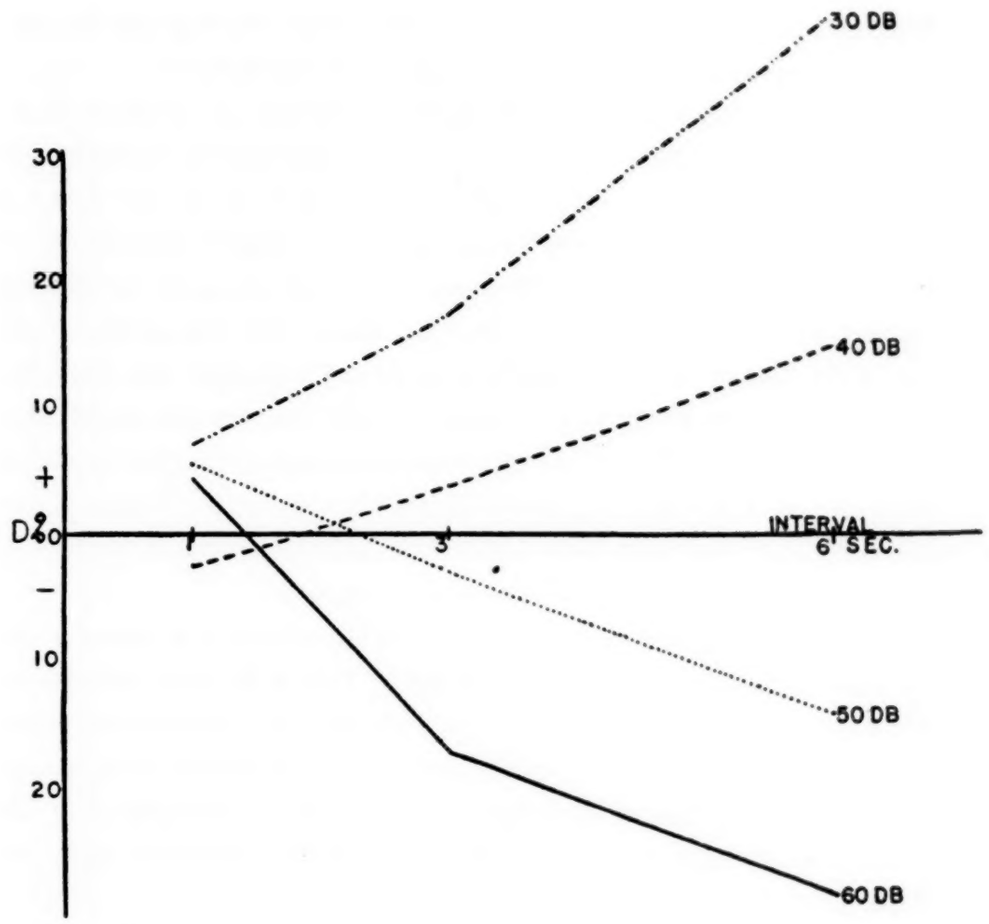


Fig. 2. Time-error values at four different levels of intensity for three time-intervals.

confirm the general findings concerning the effects of 'contrast' conditions. With a time interval of average length (3 seconds in this case) there is a regular displacement from a positive time-error for the weakest stimulus-series (30 db) to a negative time-error for the loudest stimulus-series (60 db). This accords with the results of the investigations cited above, and is what is to be expected from the nature of the experimental conditions.

The data pertaining to the effect of a variation of the time interval are clear-cut and consistent. If the results for the 6 second interval are compared with those for the 3 second interval, it will be noted that in the case of the weaker stimulus-series (30 and 40 db) there is an increase in the relative under-estimation of the comparison stimuli (indicated by a positive time-error of greater magnitude), while for the

louder stimulus-series (50 and 60 db) the change is in the opposite direction of a greater negative time-error.

In the case of the shortest time interval (1 second) the time-error differences for the various intensive levels are smaller and probably less significant. There is in Fig. 2 a suggestion that with this interval the relative positions of the levels with respect to their time-errors tend to be reversed, and there is support for this in the data for the individual Os, as well as in unsystematic observations by the writer. This reversal involves the relative under-estimation of the louder stimuli and the relative over-estimation of the weaker stimuli; it is entirely consistent with previous *p*-function findings, which indicate that time-errors tend to have opposite directions at the longer and shorter intervals.

The especial significance of the results which are presented here is the fact that the *p*-functions for the different intensive levels slope in opposite directions. A further question arose as to the range of intensive differences within which this result would obtain. A supplementary experiment, attempting to ascertain this range, will be reported before a general discussion is undertaken.

II

A majority of the experiments which involve or which study the comparison judgment employ the Method of Constant Stimuli, or some variation of it. In this method each of a series of comparison stimuli is compared with a single standard, which has usually a value approximately equal to the mid-point of the comparison series. Longer and more tedious, and consequently less used, is the Method of Paired Comparisons, in which each comparison stimulus is compared with itself and with every other stimulus. This method provides a wider range of equal pairs and a larger number of step-differences in both directions, with more than a single case of a stimulus-difference of a given magnitude and direction.

A series of five stimuli was subjected to the Method of Paired Comparisons. The stimuli, auditory intensities, were at a level 40 db above an approximated RL, and the step-

differences were 1 db in magnitude. In the interest of convenience the five stimuli are designated by the capitals *A*, *B*, *C*, *D*, and *E*, in order of decreasing magnitude, *A* being the loudest and *E* the weakest. There were twenty-five comparison pairs, *A–A*, *A–B*, ··· *A–E*, *B–A*, *B–B*, ··· *E–D*, *E–E*.

Two observers, Mr. G. L. Follansbee and Mrs. E. B. Needham, made twenty judgments upon each of the comparison pairs with the stimuli separated by an interval of 1 second, and twenty judgments with an interval of 6 seconds. Each experimental session included the use of both time intervals. The various comparison pairs were presented in haphazard order, and the observers were instructed to report the second member of each pair as 'Louder,' 'Equal,' or 'Weaker' than the first.

The first six columns of Table 3 present the frequencies of the three categories of judgment upon the twenty-five pairs at each of the two intervals. The data cited are the totals for the two observers; it should be remarked that all the significant features of the individual results are preserved in the totals.

The subsequent two columns of Table 3 are obtained by transforming the obtained frequencies of judgment for each comparison pair at the two intervals into a measure of the relative preponderance of the one or the other category of judgment, 'Louder' or 'Weaker.' The measure is obtained by applying the formula for the D percent (above) to the judgments upon each comparison pair, as well as to the judgments from the series as a whole; the measure is positive when judgments 'Weaker' preponderate, and negative when judgments 'Louder' are in the majority.

The ninth and final column of Table 3 indicates the amount and direction of the change in the measure of preponderance when the time interval is lengthened from 1 second to 6 seconds. A positive value in this column indicates an increase in the relative number of judgments 'Weaker,' while a negative value means a greater number of judgments 'Louder.'

Upon the basis of the values of the D percent for the total series of stimuli the results for the Method of Paired Compari-

TABLE 3

DATA FOR THE METHOD OF PAIRED COMPARISONS WITH TWO TIME INTERVALS

Comparison Pair	Distribution of Judgments						Measure of Preponderance		Change: 1 sec to 6 sec
	1 second			6 seconds					
	L	E	W	L	E	W	1 second	6 seconds	
A–A	18	3	19	37	1	2	+ 2.5	− 87.5	− 90.0
A–B	11	16	13	16	13	11	+ 5.0	− 12.5	− 17.5
A–C	4	12	24	2	15	23	+ 50.0	+ 52.5	+ 2.5
A–D	1	2	37	0	2	38	+ 90.0	+ 90.0	0
A–E	0	0	40	0	0	40	+100.0	+100.0	0
B–A	29	11	0	38	2	0	− 72.5	− 90.0	− 17.5
B–B	12	21	7	24	13	3	− 12.5	− 52.5	− 40.0
B–C	4	17	19	8	11	21	+ 37.5	+ 32.5	− 5.0
B–D	0	2	38	1	3	36	+ 90.0	+ 87.5	− 2.5
B–E	0	0	40	0	0	40	+100.0	+100.0	0
C–A	32	6	2	40	0	0	− 75.0	−100.0	− 25.0
C–B	21	13	6	30	7	3	− 37.5	− 67.5	− 30.0
C–C	7	19	14	16	9	15	+ 17.5	− 2.5	− 20.0
C–D	3	12	25	4	10	26	+ 55.0	+ 55.0	0
C–E	0	6	34	0	1	39	+ 70.0	+ 95.0	+ 25.0
D–A	36	4	0	40	0	0	− 80.0	−100.0	− 20.0
D–B	28	8	4	30	6	4	− 60.0	− 65.0	− 5.0
D–C	14	18	8	22	10	8	− 15.0	− 35.0	− 20.0
D–D	8	14	18	1	17	22	+ 25.0	+ 52.5	+ 27.5
D–E	2	13	35	0	2	38	+ 82.5	+ 90.0	+ 7.5
E–A	39	1	0	40	0	0	− 95.0	−100.0	− 5.0
E–B	34	5	1	40	0	0	− 82.5	−100.0	− 17.5
E–C	25	11	4	30	6	4	− 52.5	− 65.0	− 12.5
E–D	10	22	8	13	11	16	− 5.0	+ 7.5	+ 12.5
E–E	12	17	11	1	13	26	− 2.5	+ 62.5	+ 65.0
Totals	350	253	397	433	152	415	+ 4.7	− 1.8	− 6.5

sons are in agreement with other *p*-function results; the D percent is positive at the 1 second interval, negative when the interval is 6 seconds. This relationship held for both of the individual observers. However, an inspection of the data for the different comparison pairs reveals that this displacement of the time-error values is not distributed equally through the range of stimulus-pairs. Indeed, the displacement which occurs is in opposite directions at the opposite extremes of the series. In the comparison of the loudest stimulus with itself (*A–A*) there is at the 6 second interval an increased number of judgments 'Louder'; for the weakest stimulus-pair (*E–E*) the 6 second interval is marked by a relatively larger number of judgments 'Weaker.' The same opposition, less in amount, is present for the *B–B* and *D–D*

pairs, while the index of displacement for the *C–C* pair occupies an intermediate position. Comparable results appear throughout Table 3. Thus, for the *B–A* pair the lengthening of the interval introduces a greater number of judgments 'Louder,' while in the case of the *E–D* comparison there is a greater number of judgments 'Weaker' when the longer interval is employed, although the step-difference is the same in both instances (+ 1 db).

There is a large number of factors which condition an analysis of the data in Table 3. The opportunity or possibility of a displacement is different for different comparison pairs. However, no discussion is necessary to validate the conclusion that within the Method of Paired Comparisons as employed here a change of the time interval is accompanied by opposed effects in the judgments which are essentially identical with those which were noted in the data of the preceding section, that the comparison pairs *A–A* and *E–E* function as do the different intensive levels, and that they may be treated as such.

III

It is reasonable to state that the dissimilarity of the time-error measures of a number of inter-acting intensive levels is to be attributed to the 'pre-disposing' effects of preceding stimulus-presentations. The data of the present experiments, which indicate that this dissimilarity becomes greater with a longer time interval, suggest an extension of this generalization: *as the time interval separating the members of a pair of comparison stimuli is lengthened, the 'pre-disposing effects' of the preceding stimulation become more and more effective.*

A number of theoretical implications derive from this experimentally determined generalization. In the results of an earlier experiment dealing with the effects of repetition upon the *p*-function [10] the writer recognized the possibility of characterizing the first stages of the *p*-function as more similar to a relative judgment in the strict sense, while in the later part

[10] J. G. Needham, The time-error as a function of continued experimentation, *Amer. Jour. Psychol.*, 1935, **47**, in press.

of the *p*-function the judgment bore more resemblance to an absolute judgment. This statement substitutes the implication that as the time interval lengthens the standard stimulus becomes less and less effective as a basis for judgment for the doctrine of a 'fading trace.' The data of the present experiments seem to involve similar characteristics; as the time interval becomes longer, the various comparison stimuli are judged less upon the basis of their respective standards, and more by reference to the total range of preceding stimulation.[11]

It is hardly necessary to point out that the present data limit the validity of any theory which treats the time-error as attributable entirely to a fading or sinking of the trace of the standard stimulus. If this was the only factor involved, it would be imperative that the traces of a number of standard stimuli change in the same direction, in which case the various *p*-function curves would be mutually parallel.

The concept of assimilation, as advanced by Lauenstein and by Woodrow, is much more applicable to the findings reported here. Indeed, the curves in Fig. 2 are extremely similar in form to those presented by Lauenstein,[12] who determined *p*-functions when the pairs of comparison stimuli were presented upon a stimulus background which was either much more or much less intense than the stimuli themselves. Lauenstein argued that his results were determined by the assimilation of the trace of the standard stimulus to the background. The present results conform to the same generalization if the assimilative process is assumed to involve a displacement of the effective value of the standard stimulus in the direction of the average level of the several intensive levels. The direction of the assimilation is opposite at the extremes

[11] It is of interest to relate this distinction between relative and absolute judgment to the fact that the 'pre-disposing effects' appear to be equally strong in the two parts of the experiment, although the respective ranges of intensive levels certainly are not of the same order of magnitude. Schumann pointed out that the so-called 'absolute impression' may appear with equal vigor within any range whatever. "The less different from one another the single examples of a class of objects are, the more precise is the adjustment (*Anpassung*) to the middle size, and the smaller are the deviations which are needed to call forth an absolute impression." F. Schumann, Beiträge zur Analyse der Gesichtswahrnehmungen, III, *Zsch. für Psychol.*, 1902, **30**, 281.

[12] *Loc. cit.*, 160.

of the intensive scale; the assimilative effect is more marked the longer the time interval; and on the basis of these premises it follows that the *p*-function curves for a number of intensive levels should diverge, which is exactly what is indicated in Fig. 2.

However favorable the present results may be for a theory of assimilation, a considerable degree of caution is urgent. The general findings concerning the various 'pre-disposing' effects which have been cited are as harmonious with theories of psychological contrast as they are with the concept of assimilation, and it is just as logical to argue for a contrast-increment as the time interval becomes longer as it is to assume an increased assimilative effect. Moreover, a most perplexing difficulty for the principle of assimilation, as well as for theories of contrast, is to be found in the available data concerning the effects of interpolation. When a discrete stimulus is presented between the presentations of two comparison stimuli, it seems reasonable that its 'pre-disposing' effects would be comparable to those discussed above. From a comparison of data for judgments with and without interpolated stimuli, Pratt [13] concluded that assimilation might be demonstrated in the case of interpolation, but that the theory failed to hold with the inclusion of the data for the comparison judgment under normal conditions. The present writer has obtained evidence which indicates that the course of the *p*-function under interpolation conditions does not correspond to that obtained with other 'pre-disposing' conditions.[14]

Despite such theoretical involvements it is certain that the present results constitute a consistent addition to the knowledge of time-error phenomena. A judgment is a complex affair, and its determinants may be numerous. But whether the eventual decision favors a theory of assimilation, of contrast, or of something else, the present data supplement both the knowledge of and the significance of the *p*-function of the time-error.

[13] C. C. Pratt, The time-error in psychophysical judgments, *Amer. Jour. Psychol.*, 1933, **45**, 292–297.

[14] J. G. Needham, Interpolation effects at different time intervals, JOUR. EXPER. PSYCHOL., 1935, **18**, in press.

Fernberger [15] has evaluated the interdependence of judgments within the stimulus series which is involved in the Method of Constant Stimuli. The results provide definite evidence of 'pre-disposing' effects entirely comparable to those discussed above. From this point it should be reasonable to proceed not only to a further understanding of the *p*-function of the time-error under normal experimental conditions, but also to a thorough interpretation of the phenomena and mechanism of judgment in general.

With respect to the wider significance of these experimental results, it may be recalled that the comparison judgment was at one time considered to be an instance of general memory functions, the *p*-function being referred to alterations (fading, and so forth) in the memory image of the standard stimulus.[16] The discussion above (p. 540) pointed out that the concept of a sinking trace was not applicable to the experimental results, and that there is as yet little to choose among several factors which may participate in the phenomena of the *p*-function. This is in accord with the general systematic tendency to discard the concept of memory as a deterioration, and to consider it rather as a dynamic process in which there occur various changes which pertain to the organization of memory traces. Thus, for example, Gibson, [17] and Carmichael, Hogan and Walter,[18] have succeeded in the analysis of memory changes which are involved in the reproduction of visually perceived forms; indeed, Wulf, who also investigated such changes, proposed in explanation a principle of 'assimilativer Wahrnehmung,' [19] which is the same concept adduced by Lauenstein to account for certain time-error phenomena.

[15] S. W. Fernberger, Interdependence of judgments for the method of constant stimuli, Jour. Exper. Psychol., 1920, **3**, 126–150.

[16] J. G. Needham, The time-error in comparison judgments, *Psychol. Bull.*, 1934, **31**, 229–243.

[17] J. J. Gibson, The reproduction of visually perceived forms, Jour. Exper. Psychol., 1928, **12**, 1–39.

[18] L. Carmichael, Hogan, and Walter, Effect of language on the reproduction of visually perceived form, Jour. Exper. Psychol., 1932, **15**, 73–86.

[19] F. Wulf, Über die Veränderung von Vorstellungen (Gedächtnis und Gestalt), *Psychol. Forsch.*, 1921, **1**, 367.

In general, then, subsequent treatment of the problems of the time-error should follow the trend of the discussion of other memory phenomena in emphasizing the concept of a temporal course of reorganization of previous experience, and should attempt to analyze the forms and conditions of this reorganization.

Summary

1. The judgment of comparison series at several interspersed intensive levels results in a relative over-estimation of the louder stimuli and a relative under-estimation of the weaker stimuli.

2. With a longer time interval between the comparison stimuli the differences in the relative estimation-errors are magnified.

3. Comparable effects appear within the limits of the Method of Paired Comparisons, employed in its complete form.

4. It is suggested that a concept of reorganization is a consistent basis for discussion in the explanation of changes in judgments with a variation of the time interval.

(Manuscript received August 7, 1934)

SIMULTANEOUS LIFTING OF EQUALLY HEAVY WEIGHTS BY BOTH RIGHT AND LEFT HANDS

BY NAI-CHANG SHEN[1]

Psychological Laboratory, Tsing Hua University, Peiping, China

INTRODUCTION

The usual procedure in all lifting weight experiments is to have the subject lift successively the weights to be judged. Generally, the 'standard' weight is lifted first and then the 'comparison' weight. In some cases, a standard weight is followed by several comparison weights (**3**, 5). Anyway, the two kinds of weight, the standard and the comparison, have perhaps never been lifted simultaneously. This is natural, because only one hand of the subject is ordinarily used. In such a procedure, the factor of temporal error might have operated in one way or the other to affect the results. It is obvious that temporal error could be eliminated by simultaneous lifting. If the ordinary distinction between 'standard' and 'comparison' is purposely discarded and the direction of judgment is not specified, simultaneous lifting of two weights may give us some insight into the subject's type of judgment tendency.

But as soon as simultaneous lifting is introduced, both hands of the subject have to be used, although it is not impossible to lift two weights with two different fingers of the same hand. Temporal error being eliminated, spatial error must also be controlled by arranging the total number of weight pairs in such a way that there is an equal number of times for both right and left hands to lift heavier (or lighter) weights. However, the effect of the spatial error will be

[1] The writer is indebted to Professor Siegen K. Chou for his untiring interest in the experiment, both in criticizing and in suggesting ways of attack in the course of the study. Constant encouragement from him has been a source of inspiration for work. The writer wishes to express his sincere thanks to all the ten subjects who willingly lent their valuable time and patiently went through the several periods of experiment.

slight in the present case, since the experiment aims to study the lifting of equally heavy weights. In fact, the present experiment may be considered as an attempt to find the very effect of the spatial error due to the employment of two hands in lifting equally heavy weights.

Historically there is no antecedent work, so far as the author is aware, that is exactly like the present study. An approximation to our experiment is found in Arons and Irwin's investigation (1), in which they used equally heavy weights but successive lifting to determine the type of psychophysical judgment. They are probably the first ones to use equally heavy weights in psychophysical work. It is really interesting to note the fact that all past investigators were accustomed to use unequally heavy weights and neglected the possibility of digging deeper into the psychophysics of lifting equally heavy weights, in spite of the fact that the watchword of psychophysics has been the study of experiences of both identity *and* difference!

Our procedure differs, however, from Arons and Irwin's in one important aspect. In their study, the equally heavy standard and comparison weights were lifted successively by the same hand. In our experiment to be described presently, the two equally heavy weights were lifted by both right and left hands at the same time. The distinction between 'standard' and 'comparison' disappeared altogether in our case with simultaneous lifting, the direction of judgment being left to the subject's 'free will.'

Problem

Our problem is this: when both right and left hands are presented with equally heavy weights, would one of the hands, the preferred or the non-preferred, tend constantly to pass more judgments in one category—'lighter,' 'equal,' or 'heavier'—than the other?

If it does, (1) would this fact of passing more judgments in one definite category be something more than chance occurrence?, and (2) what determines the passing of more judgments by one hand in one category than by the other?

Method and Procedure

Altogether we used ten adult subjects, of whom five are major students of the Department of Psychology and have had some training in laboratory observation. One subject is an assistant in the Department. The other four are undergraduates from other departments and have never had any experience with psychological experimentation. Two of the major students are girls.

All the subjects were not aware of the purpose of the study and were asked only to judge the relative weight of any pair of boxes that stood in front of them.

The weights used were black, cylindrical, wooden boxes, $2\frac{1}{2}$ in. in height and $2\frac{1}{2}$ in. in diameter. The inside was filled with shoots and paraffin to the desired weight. Five kinds of weight were constructed, namely, 88, 94, 100, 106, and 112 grams. There are four boxes for each of the four weights 88, 94, 106, and 112 and ten boxes for the 100 g weight. We purposely made more boxes than necessary in order to avoid subjects' possible sophistication in recognizing the particular weight or weights, which may attach to a few boxes in usage. Whenever the same weight was used in different trials, the boxes for that same weight were used in rotation.

In presenting the weights, every pair of boxes was pushed to the front of the subject by means of a carrier containing them side-by-side about one inch apart. The subject's hands had been previously adjusted in a convenient position ready to lift this pair of boxes.

When the subject came for the first sitting, he was simply told that he was going to be presented with a pair of boxes at a time, and that his task consisted of lifting them up and judging their relative weight. Then an instruction sheet was given to him, which reads:

> "I am going to present to you two boxes. You are to lift up these two boxes, one in each hand, and to judge their relative weight. As soon as you have reached a decision, please report your judgment to me. In reporting, you may use any of these three terms: namely, '*heavier*,' '*lighter*,' or '*equal*,' and tell me which box, the one in your right hand or the one in your left hand, you are referring to. For example, if you feel the right hand box is heavier, you may report 'right heavier'; if you feel the left hand box is lighter, you may report 'left lighter'; however, if you feel the two boxes are equal in weight, please say 'equal.'
>
> "One thing you have to bear in mind is that in weighing the boxes, please judge only by the felt weight in the hand, not by anything else.
>
> "We want to know, in this experiment, how accurately you can judge the relative weight of the two boxes. Some pairs of the boxes are equal in weight, while some others are unequal in weight. But the weights of the different pairs of boxes do not follow any definite sequence from trial to trial. So it is impossible for you to guess the weight of different boxes in different trials. Therefore, please try your best to judge every time as accurately as you can."

After the subject finished reading the instruction and had no question about the task that was required of him, the experimenter showed him the way of lifting the pair of boxes with the thumb and the middle finger of each hand, holding the box at its middle. At this time three points were stressed to the subject: (1) lifting up the pair of boxes with two hands at the same time, (2) lifting them up to the same height, and (3) lowering them down at the same time.

At first we decided to have each pair of boxes lifted only once, later we allowed the subject to make more than one lifting if he was not quite sure as to the difference or

identity of the boxes. This was done for the sake of certainty of judgment at the expense of uniformity of manner of lifting. We believe that this procedure does not particularly favor any one hand. The subject was especially warned to observe strictly the three points mentioned above, in case he wanted to lift the boxes more than once. In the actual execution of lifting, our subjects all lifted the pair of boxes more than once in most of the trials.

Starting from the second daily sitting, the experimenter merely repeated to the subject, at the beginning of the experiment, the three terms to be used in reporting his judgment, namely, 'lighter,' 'heavier,' and 'equal.' Then the subject had, as a preliminary training, a few liftings of differently paired boxes before starting on the regular lifting of the paired weights. This preliminary training consists of lifting ten pairs in the first sitting and five pairs in each of the four later sittings. Altogether each subject went through five sittings distributed over five days, each at the same hour. Each sitting consists of lifting 100 pairs of boxes, thus yielding 100 judgments. A three-minute rest was allowed after finishing fifty pairs. The preliminary liftings before the formal start of each sitting were not included in the data, but the number of correct judgments was told to the subject. The results from the 100 regular judgments in each sitting were, however, kept out of the subject's knowledge.

Since our aim of the study is to investigate simultaneous lifting of equally heavy weights by both right and left hands, we arbitrarily chose the 100 g boxes to form the equally heavy pairs. In each sitting there were 60 equally heavy pairs and 40 unequally heavy pairs. Altogether we have twenty possible combinations for the unequally heavy pairs, which were formed according to the following arrangement in Table I.

TABLE I

FORMATION OF UNEQUALLY HEAVY PAIRS (IN GRAMS)

R Hand–L Hand	R–L	R–L	R–L
88– 94	88–100	88–106	88–112
	94–100	94–106	94–112
112– 88		100–106	100–112
112– 94	106– 88		106–112
112–100	106– 94	100– 88	
112–106	106–100	100– 94	94– 88

Since there were twenty combinations, each combination was presented twice in each sitting. The order of appearance, in each sitting, of these unequally heavy pairs, as well as those equally heavy ones, was by chance. We prepared five different schedules according to random order for the five sittings. Each schedule contained the order of the 100 pairs, divided into two halves. The order of following the five schedules by the different subjects is shown in Table II.

The insertion of unequally heavy pairs among the equally heavy ones (only the latter are our objective for this investigation) is to avoid possible suggestion that may result from constantly giving the 100 g boxes.

RESULTS

Since there are 60 pairs of equally heavy weights and 40 pairs of unequally heavy ones in each sitting, and since, with the unequally heavy pairs, the number of times for both right

TABLE II
ORDER OF FOLLOWING THE SCHEDULES

Subject	Schedules Followed Sittings					Which Half of Each Schedule Followed First
	1	2	3	4	5	
1	*A*	*B*	*C*	*D*	*E*	First half
2	*B*	*C*	*D*	*E*	*A*	" "
3	*C*	*D*	*E*	*A*	*B*	" "
4	*D*	*E*	*A*	*B*	*C*	" "
5	*E*	*A*	*B*	*C*	*D*	" "
6	*A*	*B*	*C*	*D*	*E*	Second half
7	*B*	*C*	*D*	*E*	*A*	" "
8	*C*	*D*	*E*	*A*	*B*	" "
9	*D*	*E*	*A*	*B*	*C*	" "
10	*E*	*A*	*B*	*C*	*D*	" "

and left hands to receive heavier (or lighter) weight is equal, it is obvious that by such an arrangement, the psychological outcome of this procedure ought to indicate an equal number of heavier (or light) judgments with both right and left hands, if there is no intrinsic difference between the two hands.

Frequency of All Judgments

Table III summarizes the frequency of various types of judgments by each subject in each of the five sittings. Let us first consider the first nine subjects 1–9. It is interesting to note that every one of them judged in terms of 'heavier,' although the experimenter called the subject's attention, at the beginning of every sitting, to all the three terms to be used in judging, namely, 'lighter,' 'heavier,' and 'equal.'

Subjects 1, 3, 7, 8, and 9 tended to give many more RH (right heavier) judgments and subjects 2, 4, 5, and 6, many more LH (left heavier) judgments. This tendency is consistent throughout all the five sittings for all subjects, except in the third sitting of subject 2 and in the fourth and fifth sittings of subject 8. In subject 2, the inversion may be just accidental; but in subject 8, he was changing definitely from one type of judgment (more RH's than LH's) to another (more LH's than RH's). It is to be noted that the latter subject, when questioned after the whole experiment was over as to the cause of his change of judgment, admitted that he was

TABLE III

FREQUENCY OF JUDGMENTS PASSED TO THE 300 EQUALLY HEAVY AND THE 200 UNEQUALLY HEAVY PAIRS OF WEIGHTS

Sitting	Types of Judgment	Subjects									
		1	2	3	4	5	6	7	8	9	10
1st.....	RH	50	12	51	21	34	30	46	32	57	18
	=	30	24	24	27	23	13	28	41	22	23
	LH	20	64	25	52	43	57	26	27	21	51
	LL										8
2d.....	RH	51	27	66	24	27	50	48	52	57	0
	=	34	42	5	30	7	5	26	31	21	16
	LH	15	31	29	46	66	45	26	17	22	55
	LL										29
3d.....	RH	53	64	45	26	33	40	79	50	62	1
	=	25	13	35	30	10	6	15	37	16	23
	LH	22	23	20	44	57	54	6	13	22	47
	LL										29
4th.....	RH	38	36	42	22	25	43	75	22	42	0
	=	30	18	45	27	9	0	16	36	40	16
	LH	32	46	13	51	66	57	9	42	18	42
	LL										42
5th.....	RH	56	18	44	29	13	37	70	21	56	0
	=	21	15	33	26	4	8	8	36	31	16
	LH	23	67	23	45	83	55	22	43	13	33
	LL										51
Total....	RH	248	157	248	122	132	200	318	177	274	19
	=	140	112	142	140	53	32	93	181	130	94
	LH	112	231	110	238	315	268	89	142	96	228
	LL										159

RH—'right heavier'; LH—'left heavier'; LL—'left lighter'; =—'equal.'

very absent-minded during the last two sittings, because he had a quarrel with somebody.

At the extreme right of Table III appear the results from subject 10. At a glance, her results were entirely different from those of all the others. She judged in terms of 'lighter,' a fact which never happened to the other nine subjects. This is due to the fact that she had the habit of judging the relative weight of any pair of boxes in terms of the one in her left hand, as she told the experimenter at the close of the last sitting. If the box in the left hand was felt heavier, she reported 'LH'; if contrary, of course 'LL.' The eighteen RH judgments in the first sitting were passed at a time when she took unusual attention and care in lifting the paired boxes. The only one

RH judgment in the third sitting was perhaps accidental. On the whole, her results are not so different from others' results. She gave many more LH's than LL's, in spite of the equal number of liftings of heavier and lighter weights for both hands. It should be remembered that in later treatment of the results, we considered, for convenience, the LL judgments as simply RH ones, except being reported in a different manner. It is extremely significant to note the fact that this tendency of giving many more LH than LL judgments seems to be decreasing more and more as the subject passed through the later sittings, with complete reversal of the relative frequencies in the last sitting. In order to bring out this gradual change in judgment tendency, we shall treat this subject's results of the first three sittings separately from the second two sittings.

Frequency of Judgments of Equally Heavy Pairs

The above discussions are confined to the results of both equally heavy and unequally heavy pairs of weights. Since the purpose of this study is to analyze the judgments passed to the equally heavy pairs of weights, we therefore make Table IV to show the judgments passed to the equally heavy pairs of weights.

Arons and Irwin have clearly demonstrated that physically equally heavy pairs are not psychologically equally heavy. In our experiment, only one of the ten subjects (subject 8) gave more 'equal' than other types of judgment to physically equally heavy weights; but, even here, the equal judgments amount to only 44 percent! This is, however, nothing new nor striking. The significant result is that, almost all subjects tended to pass 'heavier' judgments preponderantly more with one definite hand. In the case of subjects 1, 3, 7, and 9, many more RH than LH judgments were passed; and the reverse is true with subjects 2, 4, 5, 6, and 10. Would it be likely that 'heavier,' being the so-called 'positive quality' (4), was disproportionately favored, and hence the preponderant proportion of heavier judgments? This finding is especially worth noticing in view of the fact that the weights

presented to the two hands are equal in heaviness. Granted that physically equally heavy weights are not entirely psychologically equally heavy, it is perhaps reasonable to assume that, even if the subjects have the tendency to judge always in terms of 'heavier,' other things being equal, the relative proportions of RH and LH judgments may be still the same or nearly the same. But none of our subjects gave equal proportions of RH and LH judgments as shown in the last row of Table IV. This inequality in the proportions of RH

TABLE IV

FREQUENCY OF JUDGMENTS PASSED ONLY TO THE 300 EQUALLY HEAVY PAIRS OF WEIGHTS

Sitting	Types of Judgment	Subjects									
		1	2	3	4	5	6	7	8	9	10
1st.....	RH	29	3	30	12	1	16	26	17	39	9
	=	25	17	20	18	1	10	23	31	14	15
	LH	6	40	10	30	58	34	11	12	7	33
	LL										3
2d.....	RH	31	14	42	11	19	28	30	31	33	0
	=	24	32	3	22	16	5	14	24	16	14
	LH	5	14	15	27	25	27	16	5	11	35
	LL										11
3d.....	RH	33	39	28	10	9	22	52	30	40	0
	=	20	11	23	24	4	6	7	26	8	18
	LH	7	10	9	26	47	32	1	4	12	29
	LL										13
4th.....	RH	19	18	27	8	16	25	45	8	20	0
	=	25	14	30	21	7	0	12	23	34	12
	LH	16	28	3	31	37	35	3	29	6	24
	LL										24
5th.....	RH	36	6	25	14	9	20	46	8	37	0
	=	14	9	28	21	9	7	6	27	20	12
	LH	10	45	7	25	42	33	8	25	3	16
	LL										32
Total....	RH	148	80	152	55	54	111	199	94	169	9
	=	108	83	104	106	37	28	62	131	92	71
	LH	44	137	44	139	209	161	39	75	39	137
	LL										83

RH—'right heavier'; LH—'left heavier'; LL—'left lighter'; =—'equal.'

and LH judgments is summarized in Table V, which gives also the ratio between the RH and LH judgments.

The frequency of RH and LH judgments is not only unequal but the inequality is consistently greater in judging the

TABLE V

PROPORTIONS OF 'RIGHT HEAVIER' (RH) AND 'LEFT HEAVIER' (LH) JUDGMENTS

Subject	Judgments on the 300 '100–100' Pairs			Judgments on the Total 500 Pairs		
	RH		LH	RH		LH
1	148	(3.4 : 1)	44	248	(2.2 : 1)	112
2	80	(1 : 1.7)	137	157	(1 : 1.5)	231
3	152	(3.5 : 1)	44	248	(2.2 : 1)	110
4	55	(1 : 2.5)	139	112	(1 : 2.1)	238
5	54	(1 : 3.9)	209	132	(1 : 2.4)	315
6	111	(1 : 1.5)	161	200	(1 : 1.3)	268
7	199	(5.1 : 1)	39	318	(3.6 : 1)	89
8	94	(1.3 : 1)	75	177	(1.2 : 1)	142
9	169	(4.3 : 1)	39	274	(2.9 : 1)	96
10	92*	(1 : 1.5)	137	178*	(1 : 1.3)	228

* Including the LL judgments.

300 equally heavy pairs than in judging the total 500 pairs of both equally heavy and unequally heavy weights. This difference would have been even larger if we should compare separately the 300 equally heavy pairs with the 200 unequally heavy pairs. It must be pointed that in no sitting was there any inequality in the number of times for the right hand or the left hand to experience physically heavier or lighter weights.

Reliability of Judgments

Now let us turn our attention to the reliability of the above differences. Since no subject gave 'lighter' judgments (subject 10 is evidently an exception), there are only three possible types of judgments, namely, RH, LH, and =. We have found that every subject gave many more RH's (than LH's) or more LH's (than RH's). Is this merely the result of chance variation? If the subject should judge the physically equally heavy pairs merely by chance, without being affected by any particular factor, the probability for each one of the three categories of judgments (RH, LH, and =) to occur is 1/3.

We know that (2), according to Bayes' theorem, known as the theorem of Poisson, the probability of pure chance deviation lying within the limit of $p \pm L$, where

$$L = \gamma \sqrt{\frac{2pq}{N}}$$

is

$$P = \frac{2}{\sqrt{\pi}} \int_0^{\gamma} e^{-t^2} dt.$$

If we regard absolute certainty as $P = 1$ and scientific certainty or practical accuracy as $P = 0.9999779$, then $\gamma = 3$. Substituting, and performing the operations by logarithms, while remembering that, in our case,

$$p = 0.333, \quad q = 0.666, \quad N = 500 \quad \text{and} \quad \gamma = 3,$$

then,

$$L = 0.0897 \text{ or } 8.97 \text{ percent.}$$

Hence the minimum percentage of preponderance of RH or LH judgments to be significant is

$$p + L = 33.33 + 8.97 = 42.30 \text{ percent.}$$

Table VI gives the obtained and theoretical percentages

TABLE VI

OBTAINED AND THEORETICAL PERCENTAGES OF RH OR LH JUDGMENTS IN EACH SUBJECT'S 500 JUDGMENTS

Subject	Obtained		Theoretical RH, LH, or =	Minimum Significant %
	RH	LH		
1	49.6%		33.3%	42.3 (33.3+ 8.97)
3	49.6		"	"
7	63.6		"	"
8	44.7*	42.5†	"	44.8* (33.3+11.5) 47.5† (33.3+14.2)
9	54.8		"	42.3 (33.3+ 8.97)
2		46.2	"	"
4		47.6	"	"
5		63.0	"	"
6		53.6	"	"
10	46.5†	50.0*	"	44.8* (33.3+11.5) 47.5† (33.3+14.2)

* The first three sittings.
† The last two sittings.

of RH and LH judgments among 500 for each subject and also their corresponding minimum percentages to be significant. In the case of subject 10, all her LL judgments were arbitrarily turned into RH ones for convenience. The table shows that, except subjects 8 and 10, all the other obtained percentages of RH or LH judgments are well above their respective minimum significant percentages. Subject 10's results from the first three sittings are also above the minimum significant percentage. This means that the greater proportions of either RH or LH judgments are not merely chance variations. There must be some definite factor in operation that produces either more RH or LH judgments.

Judging Equally Heavy Weights by Both Hands

Similarly we treat the results of judging the physically equally heavy pairs of weights in three different ways as shown in Tables VII, VIII, and IX respectively. In Table VII, we assume that the probability of occurrence for RH, LH, and = is $\frac{1}{3}$ each. In Table VIII, the RH′ and LH′ judgments are RH + $\frac{1}{2}$ (=) and LH + $\frac{1}{2}$ (=) judgments respectively. This is in accordance with Fechner's procedure of dealing with equal judgments, in spite of the numerous objections raised

TABLE VII

OBTAINED AND THEORETICAL PERCENTAGES OF RH OR LH JUDGMENTS PASSED TO THE 300 EQUALLY HEAVY PAIRS OF WEIGHTS AND THEIR RESPECTIVE MINIMUM SIGNIFICANT PERCENTAGES (REPRESENTATION *A*)

Subject	Obtained		Theoretical RH, LH, or =	Minimum Significant %
	RH	LH		
1	49.3%		33.3%	44.8 (33.3+11.5)
3	50.7		"	"
7	66.4		"	"
8	43.3*	45.0†	"	48.2* (33.3+14.9) 51.6† (33.3+18.3)
9	56.3		"	44.8 (33.3+11.5)
2		45.7	"	"
4		46.3	"	"
5		69.7	"	"
6		53.7	"	"
10	46.7†	53.8*	"	48.2* (33.3+14.9) 51.6† (33.3+18.3)

* The first three sittings.
† The last two sittings.

TABLE VIII

OBTAINED AND THEORETICAL PERCENTAGES OF RH OR LH JUDGMENTS PASSED TO THE 300 EQUALLY HEAVY PAIRS OF WEIGHTS AND THEIR RESPECTIVE MINIMUM SIGNIFICANT PERCENTAGES (REPRESENTATION *B*)

Subject	Obtained		Theoretical RH′ or LH′	Minimum Significant %
	RH′	LH′		
1	67.4%		50%	62.2 (50+12.2)
3	68.0		"	"
7	76.7		"	"
8	65.8*	65.8†	"	65.8* (50+15.8) 69.4† (50+19.4)
9	71.7		"	62.2 (50+12.2)
2		59.5	"	"
4		64.0	"	"
5		75.8	"	"
6		58.3	"	"
10	56.7†	66.9*	"	65.8* (50+15.8) 69.4† (50+19.4)

* The first three sittings.
† The last two sittings.

TABLE IX

OBTAINED AND THEORETICAL PERCENTAGES OF RH OR LH JUDGMENTS PASSED TO THE 300 EQUALLY HEAVY PAIRS OF WEIGHTS AND THEIR RESPECTIVE MINIMUM SIGNIFICANT PERCENTAGES (REPRESENTATION *C*)

Subject	RH	LH	N (RH+LH)	Minimum Significant %
1	77.1%		192	65.3 (50+15.3)
3	77.5		196	65.2 (50+15.2)
7	83.6		272	62.9 (50+12.9)
8	78.8*	77.2†	99*	71.3* (50+21.3)
			70†	75.4† (50+25.4)
9	81.3		208	64.7 (50+14.7)
2		63.2	217	64.4 (50+14.4)
4		71.7	194	65.2 (50+15.2)
5		79.5	263	63.1 (50+13.1)
6		59.2	238	63.8 (50+13.8)
10	58.3†	73.0*	133*	68.3* (50+18.3)
			96†	71.6† (50+21.6)

* The first three sittings.
† The last two sittings.

against it. In Table IX, the equal judgments are neglected and the percentages of RH or LH judgments are calculated as fractions of the sum of both the RH and the LH judgments taken together.

A comparison of Table VII with Table VIII reveals the fact that in the latter there are more subjects whose results

are not up to the standard of required percentages. This is due to the fact that in the case of Table VII the allowed probability is 33.3 percent, while in the case of Table VIII the allowed probability is 50 percent. Splitting half of the equal judgments to both RH and LH judgments was sometimes not sufficient to raise the actual percentage of either RH or LH judgments. Table IX reveals not much change in the outcome of the results from the two preceding tables.

Frequency of Correct Judgments

The degree of accuracy of the judgments is perhaps worth giving. The subjects were allowed to make more than one lifting for a given pair of weights in order to assure greater certainty of judgment. Now we can calculate the number of correct judgments in the total 100 pairs as well as in the 60 equally heavy and the 40 unequally heavy pairs separately in each sitting. By a correct judgment we mean that the judgment agrees with physical identity or difference of the members in each pair of weights concerned. Table X shows the

TABLE X
PERCENTAGE OF CORRECT JUDGMENTS IN EACH SITTING OF 100 JUDGMENTS

Subject	Sitting					
	First	Second	Third	Fourth	Fifth	Av. %
1	57	53	53	57	44	52.8
2	42	59	44	48	39	46.4
3	53	37	49	55	60	50.8
4	44	53	58	53	53	52.2
5	46	37	43	45	33	40.1
6	42	43	40	32	42	39.8
7	53	39	31	38	39	40.0
8	61	52	51	47	57	53.6
9	46	47	35	65	48	48.2
10	43	51	51	46	45	47.2
Av. %	48.7	47.1	45.5	48.6	46.0	47.1

number of correct judgments for each subject in the total 500 judgments of the five sittings, while Table XI gives the same results treated separately for the 300 equally heavy and the 200 unequally heavy pairs of weights.

TABLE XI

PERCENTAGE OF CORRECT JUDGMENTS FOR THE 60 EQUALLY HEAVY PAIRS (E) AND THE 40 UNEQUALLY HEAVY PAIRS (U) IN EACH SITTING

Subject	1st sit.		2d sit.		3d sit.		4th sit.		5th sit.		Av. (E)	% (U)
	(E)	(U)	(E)	(U)	(E)	(U)	(E)	(U)	(E)	(U)		
1	41.7	80.0	40.0	72.5	33.3	82.5	41.7	80.0	23.3	75.0	36.0	78.0
2	28.3	62.5	53.3	67.5	18.3	82.5	23.4	85.0	15.0	75.0	27.7	74.5
3	33.3	82.5	5.0	85.0	38.4	65.0	50.0	62.5	46.7	80.0	34.7	75.0
4	30.0	65.0	36.6	77.5	40.0	85.0	35.0	80.0	35.0	80.0	35.3	77.5
5	26.6	75.0	6.7	82.5	11.7	90.0	15.0	90.0	1.7	80.0	12.3	83.5
6	16.7	80.0	8.3	95.0	10.0	85.0	0	80.0	11.7	87.5	9.3	85.5
7	38.3	75.0	23.4	62.5	11.7	60.0	20.0	65.0	10.0	82.5	20.7	69.0
8	51.6	77.5	40.0	70.0	43.3	62.5	38.3	60.0	45.0	75.0	43.6	69.0
9	23.3	80.0	26.6	77.5	13.3	67.5	56.7	77.5	33.3	70.0	30.6	74.5
10	25.0	70.0	23.4	92.5	30.0	82.5	20.0	85.0	20.0	82.5	23.7	82.5
Av. %	31.5	74.8	26.3	78.3	25.0	76.3	30.0	76.5	24.2	78.8	27.4	76.9

From Table X we see that there is no appreciable practice effect throughout the five sittings. The percentage of correct judgments in each sitting remains fairly constant for the individual subjects.

In Table XI we see that the correct judgments for the equally heavy pairs (E) are on the average only 27.4 percent and individual differences are great; but the correct judgments for the unequally heavy pairs (U) amount to 76.9 percent and individual differences are very much smaller in comparison with those for the equally heavy pairs. Likewise, there is no appreciable practice effect throughout the five sittings.

DISCUSSION

When we set out to work on the present problem, we had in mind the possible effect of handedness on the lifting weight experiment. It was thought that right or left handedness might influence the judgment of the physically equally heavy weights lifted under the same conditions by both hands at the same time, that is, right-handed people might show a different type of judgment as compared with left-handed people.

In the size-weight illusion, the factors that produce the deception of judgment seem to consist of a suggestion that big things are naturally heavier than small things of the same kind. The fact that the big box is judged as lighter than a

small box of the same weight may thus be due to the subject's exertion of a greater force in lifting the former, most probably influenced by its size. Since it might be assumed that the preferred hand (right hand of right-handed people or left hand of left-handed people) is capable of exerting greater force in lifting an equally heavy and equally sized weight than does the non-preferred hand, we started the present experiment with the expectation that all our right-handed subjects would naturally pass more RL (right lighter) or LH (left heavier) judgments than either RH (right heavier),LL (left lighter), or = (equal) judgments to the physically equally heavy pairs of weights. Apparently our expectation has not been realized! Just half of the subjects gave more RH judgments and the other half more LH judgments. We shall, therefore, center the following discussion upon the failure to verify this expectation.

It seems that the subject's 'attitude' has something to do with the types of judgment passed to the physically equally heavy weights. When each subject was asked at the close of the fifth sitting to comment separately on the experiment, all reported that the boxes were so light that they seemed to be 'floating' when lifted. Each subject was then given a 100 g box and asked to guess its absolute weight; each one underestimated the weight very much. Some said that it was only about 20 g and the largest guessed weight was not over 70 g. This agrees with the 'floating' feeling of the weight boxes.

Since the weights used range only from 88 to 112 grams, perhaps they are too light for handedness difference to manifest itself. In other words, when the weights are too light, the two hands of the same person, although having a preferrence, will be of uncertain importance in determining the kind of judgment in one way or the other, because the force needed for lifting will be very small. It may be true that the weights used here are still below the threshold or stimulus limen for handedness difference to manifest itself. Hence whether a subject should give more RH than LH judgments or more LH than RH judgments depends only on his subjective attitude

in lifting the pairs of weights. The fact that two of the subjects (8 and 10) reversed their type of judgment in the last two sittings may be due to a genuine change in subjective attitude.

Now let us recall that the subjects are of two types, one giving more LH than RH judgments and the other more RH than LH judgments. The former might be simply adopting the natural attitude toward the pairs of weights lifted, *i.e.* these subjects viewed the weights without bias and lifted them according to the natural tendency of the hands, the preferred right hand unconsciously exerting greater force and consequently yielding more LH judgments. These LH judgments, being passed because the subject continues to report the judgments in terms of the hand with the heavier weight in it, are to be regarded as another expression of the RL ones.

In the latter type of subjects who gave more RH than LH judgments, subjective attitude seems to play an important role. These subjects might be biased toward underestimating all the weight boxes because of the 'floating' feeling they gave; but, nevertheless, the preferred right hand is better qualified to appreciate the actual weight of the boxes. This general tendency of underestimation may be considered as an illusion, which is perhaps less strong with the right hand, since it has greater discriminative ability through daily usage. So, as a result of this possible difference between the two hands, more RH judgments were passed to the equally heavy pairs.

It should be noted that handedness difference may manifest itself in lifting weights under improved conditions of experimentation. Two changes in the procedure used in this experiment may favor the handedness effect: (1) to retain the original weights, but to make the dimensions of the boxes much smaller, so as to get rid of the 'floating' feeling which may eventually cause underestimation, and (2) to use much heavier weights, say 1,000 to 3,000 grams, so that there will be room enough, so to speak, for the handedness effect to function. However, before further work should be done with these modifications in mind, we have now in progress another experiment trying to determine, by the method of limits, the

relative accuracy of the two hands in simultaneous lifting of a pair of weights.

Summary and Conclusions

The purpose of this experiment is to study the judgments passed to physically equally heavy pairs of weights lifted simultaneously by both hands. One pair at a time was lifted. Unequally heavy weights were purposely inserted to avoid possible suggestions of presenting equally heavy ones only. Judgments are made in terms of 'lighter,' 'heavier,' or 'equal,' and the subject must specify which hand he is referring to.

Results were obtained from ten adult subjects. Nine of them gave judgments always with reference to the hand with the heavier weight in it. The tenth subject, differing from other subjects by giving more 'left lighter' judgments, was considered to be an exception, because her natural way was to judge the pairs of weights always with reference to that one in the left hand.

Physically equally heavy weights were judged as equal in only 27.4 percent on the average, showing that they are far from being entirely psychologically equally heavy. Furthermore, half of the subjects gave many more 'right heavier' than 'left heavier' and half of them many more 'left heavier' than 'right heavier' judgments in response to the equally heavy pairs, while theoretically both RH and LH judgments, if given at all, should be equally distributed. This preponderance of either RH or LH judgments was statistically significant in most cases.

With our technique of testing, handedness does not seem to have appreciable effect on the relative frequency of either heavier or lighter judgments of a pair of weights lifted simultaneously by both hands, because the weights used are all perhaps too light and hence below the threshold of bilateral discrimination of the two hands. It is suggested that either smaller (in size) or heavier weights may perhaps be more favorable to handedness difference in lifting weight experiments.

(Manuscript received June 18, 1934)

References

1. Arons, L., and Irwin, F. W., Equal weights and psychophysical judgments, J. Exper. Psychol., 1932, **15**, 733–751.
2. Coover, J. E., Psychical Research Monographs, No. 2, Stanford Press, 1917, 79–95. (Experiments in Psychical Research.)
3. Gahagan, L., On the absolute judgment of lifted weights, J. Exper. Psychol., 1929, **12**, 490–501.
4. Hollingworth, H. L., Experimental studies in judgment, *Arch. of Psychol.*, 1913, **4**, No. 29, ch. VI.
5. Hulin, W. S., and Frazier, L. R., Two comparison stimuli with one initial standard in lifted weight. *Amer. J. Psychol.*, 1928, **40**, 626–627.

FATIGUE, OSCILLATION, AND BLOCKS

BY ARTHUR G. BILLS[1]
The University of Chicago

A study of the successive reaction times in the continuous performance of a homogeneous mental task discloses pauses or lengthened reaction times which appear at more or less regular intervals. In such simple tasks as color-naming, form-naming, code substitution, addition, naming opposites, and the like, these pauses occur on the average about three times a minute. In a previous article, (1), the author has described these pauses under the title 'blocks,' and has pointed out that they decrease in frequency and size with practice, increase with fatigue, and vary from individual to individual, being most frequent and longest in slow workers and vice versa. For purposes of quantitative study, blocks were arbitrarily defined as pauses equivalent in length to the time of two or more average responses. This definition is not entirely arbitrary, however, for if a given subject's reaction times are plotted in the form of a frequency polygon, in which the base represents reaction time in eighths of a second, it is found that two distinct modes appear, one of which is due to the central tendency in normal reaction times, while the other shows the central tendency of the blocks. The block mode appears at a point representing approximately double the normal reaction time mode.

The relative periodicity of these blocks suggested an interpretation in terms of some hypothetical periodicity in nervous energy or capacity. A number of studies by different authors at widely different times have given evidence of periodic fluctuations in mental efficiency; among which are fluctuations in the threshold for apprehension of minimal stimuli, fluctuations in the focus of attention, fluctuations in performance level, such as those recently described in

[1] The present investigation was aided in part by a grant to the University of Chicago from the Rockefeller Foundation.

Philpotts' monograph (**2**), and fluctuations in ambiguous perspective, as well as fluctuations of much wider scope in practice curves. Spearman has given the term 'Oscillation' to these phenomena and has postulated a fundamental human trait, 'O,' which forms one of his triumvirate of human traits, 'G,' or general intelligence, 'O' or oscillation, and 'P,' or perseveration.

Since publishing his first study of blocking, the author has collected a large mass of data which seem to require interpretation in terms of more than one simultaneous rate of oscillation in neural energy, so that the resulting time curve can be accounted for only as a composite effect of the component frequencies, in the sense that a complex sound wave can be analysed into its harmonic components. A similar suggestion is offered by Philpotts in interpreting his data. The necessity of such an interpretation proves that the traditional concept of a single energy component, call it attention, or what you will, was entirely inadequate to explain the facts. The organism, in performing mental work, is not utilizing a single unitary energy, but is drawing upon a number of relatively independent energy components, each of which can be expected to have its own peculiar oscillation rate. Output is a resultant of the interaction of all of them. There is the refractory period of the sensory component, the peripheral neural and central neural components, the motor component, etc.

However, the data to be reported in this paper seem capable of being described in terms of two basic hypothetical oscillation rates.

The experiments employed about 35 subjects, advanced students in psychology. The mental work was of two sorts, color-naming and form-naming, and a work period consisted of thirty minutes of continuous responding. Two types of response were used; vocal and manual. The first consisted in pronouncing the names of the colors or forms orally as they were presented by means of a special tachistoscopic device at a rate controlled by the subject. The second involved pressing the appropriate keys on a specially prepared

keyboard. Each response was registered kymographically, and on a typewriter also, so that both time and accuracy could be computed. The instructions to the subject were to work as rapidly as possible consistent with accuracy, so that errors were at a minimum, and reaction time was the main criterion.

The kymographic records were then analysed minutely, by measuring each reaction time separately, locating the blocks, and measuring the distances between the successive blocks, both in seconds and in number of responses intervening. The data were then subjected to three different kinds of graphical analysis. First, frequency polygons were constructed in which the frequency of each length of reaction time was determined. This brought out two distinct modes, one for normal reaction times and the other for blocks. These two stood out distinctly with practically no overlapping, thus demonstrating unquestionably the reality of the block.

Next, frequency polygons were constructed in which the frequency of each inter-block interval was plotted. The inter-block intervals were computed both in seconds and in responses. In this way it was possible to determine whether there was any characteristic length of interval between blocks, or whether the blocks occurred at chance intervals. If there were any regularity or periodicity in the occurrence of the blocks, this should result in a high frequency of occurrence of some particular inter-block interval and a grouping of intervals about this as a mode. If, on the other hand, there were no periodicity at all in the recurrence of blocks, no such modal interval should appear. And if, instead, more than one modal interval appeared in the frequency polygon, it would give evidence of the presence of more than one periodicity. Thus, the same work record might show one periodicity of high frequency and another of low frequency.

Still a third type of analysis was made. Graphs were prepared in which the successive inter-block intervals were plotted against their lengths. This was done to determine whether there was any tendency for long inter-block intervals to recur periodically. Such long inter-block intervals would

indicate long stretches in which no blocks occurred, and would give indication of an over-rhythm distinct from the block rhythm itself.

Figure 1 shows the record of one subject plotted in accordance with the third method of analysis. It represents fifteen minutes of color-naming, with responses given manually. The scale points on the abscissa represent the successive intervals between blocks, while the vertical distances represent

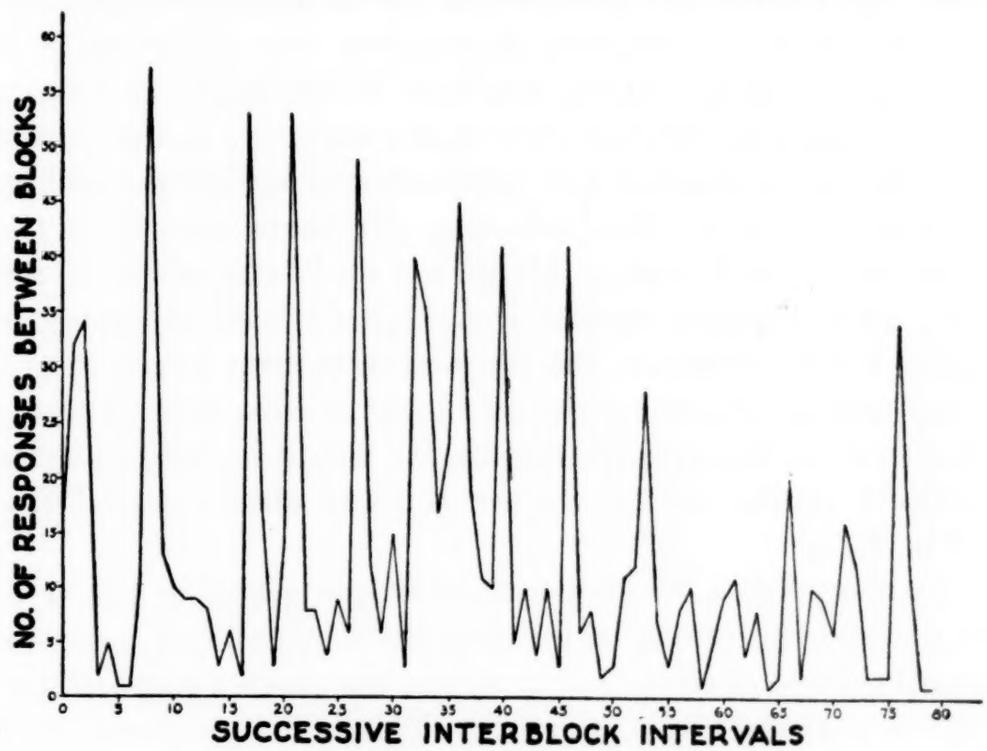


FIG. 1. Record of Subject F, showing the recurrence of a prolonged block-free interval after approximately every five blocks.

the length of the interval computed in response times. The striking characteristics of the graph are, first that about every fifth inter-block interval is prolonged to a length which is several times that of the average interval. This means, in terms of the work record, that there will be a series of blocks occurring at fairly regular intervals, say every four to six responses, and then after about four of these have occurred, there will be a long stretch free from blocks, for at least forty responses, then another group of blocks at short intervals, and then a block-free stretch, etc.

Now if we assume that the series of short inter-block intervals shows an underlying block rhythm of fairly high frequency which can be explained in terms of an energy fluctuation of similar frequency, then the prolonged inter-block intervals must indicate some second oscillation rate of much lower frequency super-imposed on the first. The blocks are assumed to occur at the troughs of the energy waves, and to indicate that the level of efficiency has momentarily fallen below the point where the subject can respond at all. A schematic diagram illustrating this relationship is given in Figure 2. At the left side of the figure we have a series of short oscillations with blocks occurring at the trough of each. This series of four blocks occurs during the trough in the longer wave. But, following this, there is a crest in the longer wave, and during this period no blocks occur at the short wave troughs because the subject's level of energy is high. When, however, the long wave returns again to the trough phase, another series of blocks occurs, and so forth. Therefore, on theoretical grounds, we would be led to expect a picture similar to the one we actually obtain empirically in Fig. 1.

Incidentally, a striking feature of the graph in Fig. 1 is the gradual shortening of the prolonged inter-block intervals toward the end of the work curve, and the greater proportional frequency of the short intervals. This is in agreement with the previously reported finding that fatigue causes an increased frequency of blocking.

Another feature which is characteristic of all the individual records is that there is a prolonged interval without blocks at the very beginning of the work record. This corresponds to the initial high point of efficiency usually found in work curves for homogeneous tasks, known as initial spurt. It suggests that initial spurt is the initial crest in the long energy wave, and that it is not an isolated phenomenon but merely an exaggerated instance of the same oscillatory tendency which persists throughout the work record. Incidentally, if the subject always begins at the crest of the large

wave, this wave must result *from* the work, and is not caused by any independent organic rhythm.

Figure 3 is a frequency diagram constructed in accordance with the second method of analysis suggested earlier in this article. It shows the number of times each different length of inter-block interval occurs, in the case of a single subject, G, who worked at color-naming and responded manually. It can be seen at once that more than one mode appears. The first mode is at an inter-block interval of one response. This

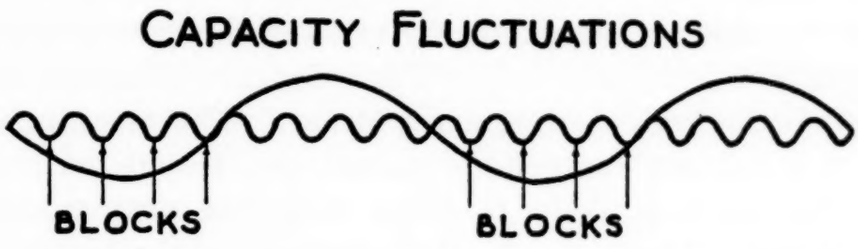


FIG. 2. Schematic diagram to represent the relation of two hypothetical energy fluctuations to the occurrence of blocks.

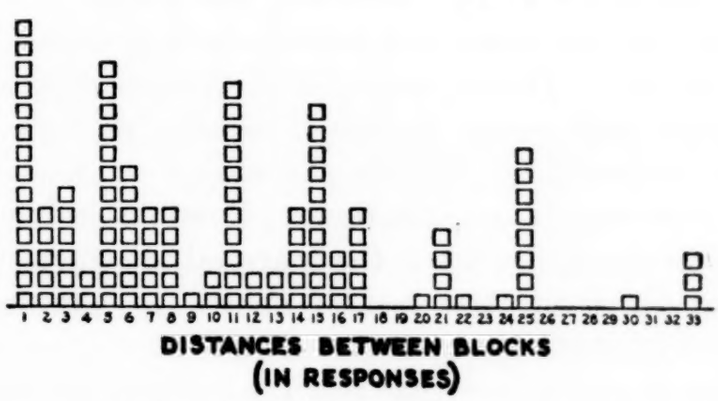


FIG. 3. Frequency diagram showing the relative frequency of occurrence of inter-block intervals of different lengths in the record of Subject G.

is characteristic for all subjects and means that two blocks occurred in immediate succession during a single trough. These cases are best considered as one long block rather than two, as far as their explanatory basis is concerned. The next mode occurs at the five point. This is a true interval and represents the recurrent trough in a wave of fairly high frequency. The next mode is at the eleven point, the next at fifteen, and thereafter no clear mode emerges until the twenty-five response interval. How can these facts be

brought in harmony with the theoretical picture of Fig. 2? Presumably there should be only two modes, one at 5 and one at 25, for example; moreover, the shorter interval should occur much more frequently, indicating blocks at the troughs of the shorter wave, while the infrequent long intervals would be due to the crest phases of the longer wave during which no blocks occur. This would leave unaccounted-for the modes at 11 and 15. But why must it be assumed that a block will invariably occur at every short-wave trough? If one trough were occasionally skipped, it would give intervals of 10 (which is approximately 11), and if two were skipped, it would give intervals of 15. The frequency diagram shows that such intervals do frequently occur. The assumption is that, if a subject successfully passes one trough without a block, no block is likely to occur until the next trough is reached, etc. Theoretically, then, the sum of the frequencies at 5, and 11, and 15, and 20, should be equal to (or less than) five times the frequency at 25. Actually this is true. But it does not account for the occasional intermediate intervals, those at 7, 8, 14, 16, etc. The explanation of these is difficult unless we recognize that many chance influences are operating in any work situation to disturb the major tendency and to produce unaccountable variations. Possible influences are variations in the difficulty of the material, distractions, etc.

In general, then, the results shown in the frequency diagram corroborate the hypothesis.

In order to make sure that this tendency is not one which is specific to one type of task or to one method of responding, the author made a comparative study of the form-naming and color-naming data, and of the vocal and manual responses. Fig. 4 shows the composite record for a group of fourteen

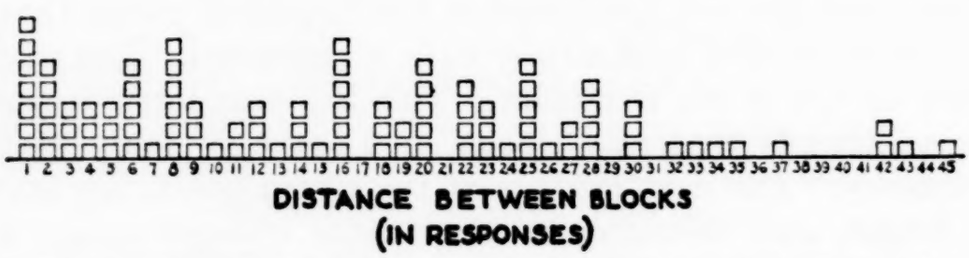


FIG. 4. Composite frequency diagram for form-naming, vocal and manual.

subjects for both manual and vocal form-naming. It is constructed in the following way: first, frequency diagrams were prepared for each of the subjects and the modes determined. Then a frequency diagram was constructed from these individual mode points. The result is a diagram showing the modes of the modes. While the intervals differ slightly from those in Fig. 3, the general appearance is the same. Modes emerge at 1, 6, 9, 16, 20 and 25. The mode at 9 is more prominent than that at 6, which is in agreement with the finding that blocking is less frequent in vocal than in manual responses.

A comparison of color-naming with form-naming can be made by referring to Fig. 5. This is the composite record

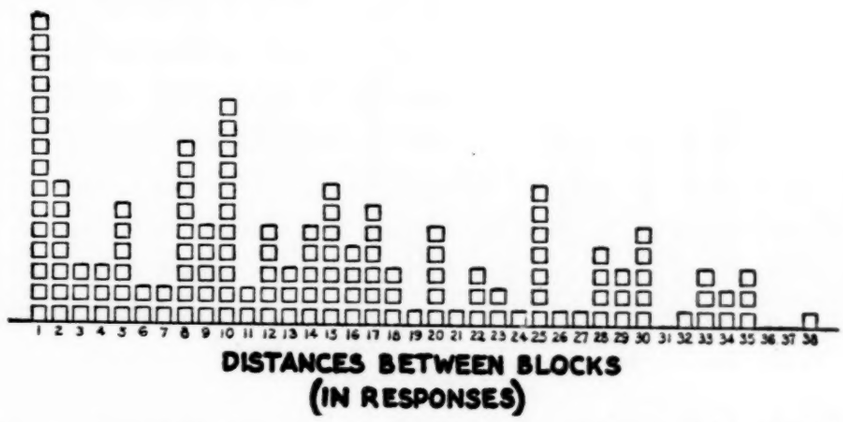


FIG. 5. Composite frequency diagram for color-naming, vocal and manual.

for a group of 20 subjects engaged in color-naming, vocal or manual, and is obtained similarly to Fig. 4. It shows prominent modes at or about 1, and 10, and 15, and a clear mode at 25. Smaller aggregations appear at other intermediate intervals.

Finally, in Fig. 6, is presented a composite record of Figs. 4 and 5. It includes the records of 34 subjects. One would scarcely expect any very clear modes to emerge from such a hodge-podge of data, but, as a matter of fact, such modes can be distinguished at 1, at 5–6, at 8–10, at 15–16, at 20 and at 25. It must be admitted that the apparent modes at 12 and 22 are unaccounted for.

A comparison of all these results points unmistakably to a

combination of two (or possibly more) rhythms underlying the blocking phenomenon. One question, as yet unanswered, is whether these rhythms are *time* rhythms or *amount* rhythms. That is, is the length of a wave measured in seconds or in number of responses? If an individual is working at a rapid rate, so that he makes more responses per unit time, will the pulsations be correspondingly more frequent, or are they temporally constant? To answer this, an analysis was made of the inter-block intervals in terms of seconds instead of responses. Frequency polygons were constructed on this basis. They do not differ materially from those of Figs. 5 and 6. But this is obviously due to the fact that the subjects'

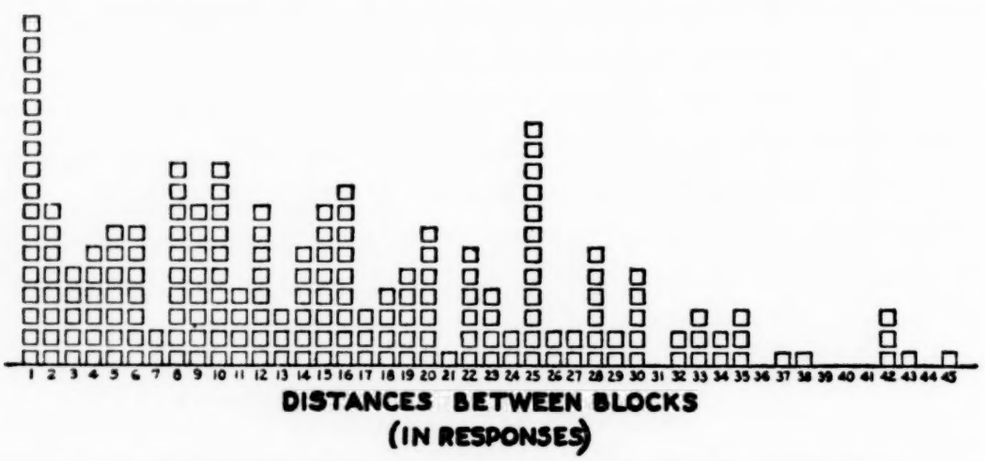


FIG. 6. Composite frequency diagram for form and color-naming both.

reaction time did not vary markedly during the work period. In a previous study, however, in which the subjects' speed was systematically varied between a low of 40 and a high of 120 responses per minute, the frequency of blocking was found to vary with the rate. That is, the blocks per minute increased in proportion to the number of responses per minute. The inter-block intervals became progressively shorter as measured by seconds. Did they remain approximately the same length as measured by responses? This can be determined from the table given below:

Responses per minute..	40	50	60	70	80	90	100	110
Blocks per minute.....	0.8	0.3	1.2	2.4	3.8	7.5	11.7	19.0
Ratio of blocks to responses.....	50	166	50	29	21	12	8.5	5.8

It appears that the inter-block intervals became shorter in terms of responses also. But these results need interpretation. The data are based on an average of a large number of subjects, not all of whom worked at all the rates. And many of the subjects had reached virtual confusion at a rate of 90 or 100. Moreover, the rates were automatically set by the speed of the exposure apparatus, whereas ordinarily a subject works at his own determined rate. The automatic control tends to reduce the blocking considerably at the lower rates. Finally, the increased frequency of blocks meant an increased frequency of the same short inter-block interval which appeared as a modal interval even at the slower rates. It did not mean an entirely new modal interval. It meant a piling up of cases around the five-mode, and fewer cases at the ten-mode, fifteen-mode, etc. The conclusion can probably be drawn safely that the fundamental rhythm, which is an *amount* rhythm rather than a *time* rhythm, remains nearly the same, but that the subject, when working at a very rapid rate, blocks at every trough, whereas, when working slowly, he passes many troughs without blocks.

In conclusion, then, the theory is advanced that the blocking phenomenon can be explained in terms of at least two fundamental periodicities in the flow of available (neural) energy, and that these periodicities depend on the amount of work done, *i.e.* the rate at which the available energy is being consumed rather than upon the amount of time which has elapsed. It is possible that, when the subject is responding very rapidly, more energy is consumed per response, so that the rhythm becomes faster as measured by responses. Why there should be two fundamental periodicities is not known, but it is reasonable in view of the fact that more than one organic component is involved in the performance, and each component may be expected to have its own characteristic exhaustion-recovery rate.

Are Blocks Resting Periods?

In a previous article (1) it was suggested that the blocking phenomenon explains the absence of large decrements in curves of mental work. It was suggested that the blocks

might be enforced resting periods which enable the organism to continue almost indefinitely without a breakdown such as muscular work curves show. In order to test the validity of this theory, an experiment was performed in which artificial rests were introduced into the continuous performance of a group of subjects, to determine whether, under these conditions, the blocks would disappear. The method was first to determine the average number of blocks per minute, and average length of blocks of each subject when working at his maximum automatic speed (consistent with accuracy) in color and form naming. Then the exposure apparatus was set to present the stimuli at this speed, with automatic pauses (of a duration equal to the length of an average block), occurring at a frequency as near as possible to the subject's average block frequency. In this way, artificial rests were substituted for the natural ones. This resulted in the almost complete disappearance of the blocks from the subjects' records. It is concluded that the artificial pauses fulfilled the same function as the natural ones, and therefore removed the need of blocks as rests. The following table shows the comparative results for the two conditions, *i.e.* with and without interpolated rests, for Subject D. First, her self-determined maximum rate for color-naming, vocal, was found to be 100 per minute. At this rate, she made an average of three blocks per minute. The machine was then set at an automatic rate of 120 per minute, at which rate she blocked, on the average, 3½ times per minute. Finally, artificial pauses of 1½ seconds duration were introduced four times per minute. In the table are given the results for five minutes work without, and five minutes work with rests.

Without Rests				With Rests			
Minute	Resp. per Min.	Blocks per Min.	Average Length in Mill.	Minute	Resp. per Min.	Blocks per Min.	Average Length in Mill.
1	111	3	2½	1	114	0	—
2	112	3	2½	2	115	1	2
3	110	3	3¼	3	120	0	—
4	112	2	4	4	118	1	1½
5	108	4	3	5	114	1	2

The total number of blocks is reduced from 15 to 3, and the size is also reduced somewhat. While this constitutes corroborative evidence that the natural blocks serve as rest periods, it is not conclusive. The rests should be used in voluntary work also, and should be spaced so as to coincide exactly with the troughs in the subject's capacity wave as empirically determined. A much more extensive exploration of this problem is planned.

(Manuscript received May 20, 1935)

References

1. Bills, A. G., Blocking: a new principle of mental fatigue. *Amer. J. Psychol.*, 1931, **43**, 230–245.
2. Philpott, S. J. F., Fluctuations in human output. *Brit. J. Psychol. Monog. Suppl.* 1932, **17**, 1–125.

THE INFLUENCE OF RELEVANCE AND BELONGING *

BY EDWARD L. THORNDIKE AND IRVING LORGE

Division of Psychology, Institute of Educational Research, Teachers College, Columbia University

Thorndike ('33) has shown that a satisfying after-effect not only strengthens the connection to which it belongs but also that it strengthens connections that are near it, even though the near connections are themselves punished. The action of a satisfying after-effect may be likened to a natural force which is directly applied to do work. Most of the force operates on the connection to which it is directed. Some of the force works backward and forward to influence connections within the field of the force's application.

It is the purpose of this experiment to ascertain the influence of satisfying connections that are not applied directly to *any one* connection. After-effects may belong definitely and clearly to the connections that they follow, or again they may be ambiguous and indefinite in their belongingness. This experiment investigates the action of after-effects of four different degrees of relevance and belongingness:

1. After-effect that belongs definitely and only to the connection it follows,
2. After-effect that may belong to any (or every) one of eight connections preceding it in time,
3. After-effect that may belong to any (or every) connection that has been made up to the introduction of the after-effect,
4. After-effect that may belong to several preceding connections.

These after-effects differ in relevance and belongingness as much as most rewards do differ. For example, after-effects

* The investigation reported here was part of a general program supported by a grant from the Carnegie Corporation.

that belong definitely and only to the preceding connection are like the teacher's 'Right' for the answer '5' to the stimulus '2 + 3'; after-effects that belong to any or all of a specified series of connections are like a teacher's '90 percent' for a spelling test, or '75 percent' for a test in Latin; after-effects that belong to any one or all connections that have been made are like the teacher's 'You're doing nicely' or 'You're improving or 'Good' for a general evaluation of performances; after-effects that belong to several connections preceding are like the teacher's 'Keep it up' for two or three correct or good performances in succession. These after-effects differ primarily in the degree of belongingness to the connection that they follow. 'Right' belongs only to the connection it follows. '75 percent' or '90 percent' belongs to all the connections of a series; 'Good' belongs to anything or everything that preceded it; and 'Keep it up' belongs to several but an indeterminate number of connections. Differences in the specificity of an after-effect are primarily a function of the belongingness of the after-effect.

This experiment investigates the potency of various after-effects of differing degrees of belongingness and relevance. The connections are followed by after-effects which varied from the definite 'Right' or 'Wrong' to the ambiguous "Here is a bonus for faithful service."

The materials of this experiment were twenty-four different series of 40 words. Each word in a series was read in a monotone. The subject responded to each word by saying a number from 1 to 5. The directions under which the subject responded were:

"This is an experiment in learning. I shall say a word. You will at once say a number from 1 to 5. Sometimes I will tell you whether you are right or wrong. Usually I will not. But I will tell you whenever you are doing specially well; and I will tell you after every eight responses how well you did in those eight. You will be paid a bonus of one cent for every two that you get right more than you would get by chance. There will be 40 words in each series, and the series will be repeated 5 times. The first time you naturally will

get the right numbers only by chance, but the second time you will do a little better, and the third time still better, and so on. Remember you are always to respond by saying 1 or 2 or 3 or 4 or 5 just as soon as I say a word."

The announcement of 'Right' or 'Wrong' was always made for the responses to 7 or 8 stimulus words in each series of forty. For the other 32 or 33 words the experimenter recorded whether the responses were correct or incorrect according to a key. No announcements of 'Right' or 'Wrong' were made for these 32 or 33; but if the subject had two responses correct in sequence, the experimenter said 'You are doing finely now' in an encouraging tone. From time to time also, but never more than three times in a single session, the experimenter handed the subject five cents in cash, saying "Here is a bonus for faithful work." Also, after the eighth, sixteenth, twenty-fourth, thirty-second and fortieth word of each series, the experimenter said "You had—out of eight right," followed by 'very poor' if 0 right, 'not very good' if 1 right, 'good' if 2 right, 'very good' if 3, 'fine' if 4, 'wonderful' if 5, 'marvelous' if 6 or 7, 'perfect' if 8. The experimenter's tonal expression also changed from sad for 0 or 1 to joyous for the better performances.

Each series was presented in the same order for five trials to each subject. At the end of each series the subject was told the total number of responses out of the 200 that were recorded as 'right' or 'correct,' and paid in cash a bonus of one-half cent for each 'Correct' or 'Right' in excess of 40.

The after-effects whose potency we seek to measure are:

(N) No announcement; the subject responds to a stimulus with a number from 1 to 5 but the experimenter does nothing, and reads the next stimulus word in a monotone.

(R) Announcing 'Right': the subject responds to a stimulus and the experimenter announces 'Right.'

(X) Announcing 'Wrong': the subject responds to a stimulus word and the experimenter announces 'Wrong.'

(F) Saying "You are doing finely": the subject responds to a stimulus word and the experimenter says "You are doing finely."

(B) Saying "Here is a bonus for faithful work." The subject responds to a stimulus and the experimenter says "Here is a bonus for faithful work."

(0, 1, 2, 3, 4, 5, 6, 7, or 8) Saying "You had—out of eight right" plus a descriptive adjective: the subject responds to a stimulus word and the experimenter makes the announcement concerning the performance with the preceding eight connections.

N, R, X, F, B, 0, 1, 2, 3, 4, 5, 6, 7, and 8 represent after-effects whose belongingness varies from the definite and unambiguous R and X, to the indefinite and equivocal B.

In Table 1, we show a sample series of forty words which were read to the subject. Next to each word is the key by which the experimenter records the response. If an asterisk appears after the number, the experimenter announces 'Right' or 'Wrong' as the subject's response is the same or different from the key. For these responses, the experimenter records R or X. If the response is the same as the key for a stimulus word which is not asterisked, the experimenter records C but makes no announcement. If the response is different from the key for a word which is not asterisked, the experimenter records the actual response but makes no announcement. If two C's occur in succession, or if the record shows a C and a R, or a R and a C in succession the experimenter says "You are doing finely" and records F after the response after which such announcement is made. At the end of each eight responses, the experimenter counts the C's and the R's and says "You had—out of eight right" plus the appropriate descriptive adjective and records the number corresponding to such an announcement after the eighth word.[1] If he announces "Here is a bonus for faithful service" and gives the subject five cents, he records B after the connection after which such announcement and bonus is given.

For any series, we have five columns corresponding to the subject's responses and the experimenter's announcements for the five trials with the series.

[1] He records 0 for no C's or R's, "1" for one C or R, "2" for two C's or R's, etc., etc.

TABLE 1

THE STIMULUS, KEY, SUBJECT'S RESPONSE, AND EXPERIMENTER'S RECORD FOR A SAMPLE TRIAL TO SHOW HOW THE VARIOUS AFTER-EFFECTS WERE APPLIED AND RECORDED

Stimulus	Key	Subject's Response	Experimenter's Record
acerbity	2	4	4
pray	3	1	1
goose	1	2	2
well being	2*	2	R
plough	4	4	C (F)
loving	3	5	5
horse	5*	4	4 X
swear	1	2	2 (2)
inept	2	1	1
alack	4	5	5
court	5	2	2
admire	5	4	4 X
youth	3	4	4
skilled	2	2	C
senator	1	5	5
dusk	4*	2	2 X (1)
adjudge	5	1	1
odd	4	5	5
insanity	3*	3	R
hip	1	2	2
attack	5*	5	R
chill	2	3	3
carbine	3	2	2
teethe	4	4	C (3)
pasture	1*	1	R
auburn	4	3	3
nut	2	2	C
above	3	5	5
onward	5	4	4 (B)
diadem	1	1	C
decision	3	3	C (F)
folly	2	4	4 (4)
banner	4*	3	3 X
war-cry	1	3	3
statue	5	1	1
recession	5	2	2
council	3*	3	R
remnant	2	5	5
piety	4*	4	R
paralysis	1	4	4 (2)

Twenty college undergraduate women took this experiment during November and December 1932. They did the twenty-four series in six hours distributed throughout a single week. Each subject was paid a basal rate for her service during the hour plus whatever additional money bonus for scores over chance and for "Here is five cents for faithful service" she earned or received.

Consider first the influence of a clearly relevant and closely belonging after-effect, the announcement of 'Right.' This causes 56 percent of repetitions in the following trial (55.81 from 1403 cases, using first occurrences only). When this announcement is accompanied by "You are doing finely" or by "You had 5 right; Wonderful!" or by "You had 6 (or 7) right; Marvellous!" or by "You had 8 right; Perfect!" the percentage is almost the same (53.49 for 301 cases of C + F; 57.14 for 14 cases of C + 5, C + 6, C + 7, and C + 8). By sheer chance the percentage of repetition would be 20. By chance plus whatever favoritisms in the use of numbers the subjects may have had, it would be somewhat but not much above 20. (On the basis of similar experiments we estimate 22 for the average subject.)

Consider next the influence of the less closely belonging 0, 1, 2, 3, 4, 5, 6, 7, 8, and F, and also of the N and X.

Since the influence of a satisfying after-effect that definitely belongs to the connection it follows usually influences not only the connection it follows but also connections near it in time or in steps, our records are analyzed to take account of the proximity of connections to an announced 'Right.' Each connection other than that of an announced 'Right' was placed in a category corresponding to the number of connections by which it precedes or follows an announced 'Right.' Table 2 presents the data for the influence of specified after-effects (N, 0 and 1, 2, 3 and 4, 5, 6, 7, and 8, F, and X) classified according to the remoteness of the connection from an announced 'Right.' Only first occurrences are used. The percentage of repetitions of the connection in the following trial gives a measure of the influence of the specified after-effect. Table 2 will also show the existence and magnitude of the 'spread' or 'scatter' phenomenon.

TABLE 2

THE POTENCY OF DIFFERENT AFTER-EFFECTS AFTER TAKING ACCOUNT OF THE INFLUENCE OF PROXIMITY TO A "RIGHT." THE FREQUENCY WITH WHICH A CONNECTION AT ITS FIRST OCCURRENCE WITH A SPECIFIED AFTER-EFFECT WAS REPEATED (S) OR REPLACED BY A DIFFERENT CONNECTION (D) AT THE NEXT TRIAL

	(N) No Announcement			(0, 1) "You had 0 or 1," etc.			(234) "You had 2, 3, or 4," etc.		
	$\frac{100S}{S+D}$	S	D	$\frac{100S}{S+D}$	S	D	$\frac{100S}{S+D}$	S	D
One step removed									
before	28.83	567	1,400	27.15	41	110	30.32	47	108
after	33.75	534	1,048	33.33	15	30	34.71	59	111
Two steps removed									
before	26.95	502	1,361	28.04	30	77	31.91	30	64
after	31.53	572	1,242	26.42	14	39	23.94	45	143
Three steps removed									
before	31.04	455	1,011	21.83	31	111	28.57	26	65
after	28.97	392	961	19.05	8	34	28.34	53	134
Four steps removed									
before	30.02	324	755	26.26	47	132	24.68	39	119
after	30.13	304	705	24.39	20	62	27.46	67	177
Five or more steps removed									
before and after	27.74	6,242	15,263	25.74	443	1,278	29.80	396	933

	(5678) "You had 5, 6, 7, or 8," etc.			(F) "You are doing finely"			(X) Announced "Wrong"		
	$\frac{100S}{S+D}$	S	D	$\frac{100S}{S+D}$	S	D	$\frac{100S}{S+D}$	S	D
One step removed									
before		3	5	32.89	25	51			
after		3	12	49.33	183	188		0	2
Two steps removed									
before		1	4	36.51	23	40		5	1
after		1	8	31.25	10	22		1	8
Three steps removed									
before		1	0	25.53	12	35	26.55	30	83
after		2	7	34.48	30	57	24.77	27	82
Four steps removed									
before		0	1	33.33	10	20	26.52	74	205
after		5	8	42.11	16	22	24.21	77	241
Five or more steps removed									
before and after	34.37	11	21	34.64	230	434	23.98	1,201	3,808

In Table 2, the influence of the after-effect may be considered unaffected by the 'spread' or 'scatter' from an announced 'Right' if the connection is five or more steps removed from the influence of such an announced 'Right.' The percentage of repetition for 'No announcement' is by this measure 27.7. Similar percentages for each of the other specified after-effects five or more steps removed from the influence of an announced 'Right' are as follows: "You had 0, 1 right, etc.," 25.7; "You had 2, 3, 4, right etc.," 29.8; "You had 5, 6, 7, or 8 right, etc.," 34.4; "You are doing finely," 34.6; an announced 'Wrong,' 24.0. The percentages of repetition for N, 0 or 1, 2, 3, or 4, 5, 6, 7, and 8, F and X four steps removed from an announced 'Right' are, in order, 30.1, 25.7, 26.3, 35.7, 38.2, and 25.3. The percentages of repetition at a remoteness of three steps are, in the same order, 30.0, 21.2, 28.4, 30.0, 31.3, and 25.7. "You are doing finely" has a marked strengthening effect. "You had 5 (or 6 or 7 or 8) right" with its accompanying encouraging adjective has also, so far as our scanty data about it can be trusted. 'Wrong' and "You had 0 (or 1) right" with its 'Very poor' or "Not very good" are less strengthening than no announcement at all. These and other percentages of repetition, when expressed as excesses over the 22 percent which would occur by chance and favoritism in the use of the numbers 1, 2, 3, 4, and 5, appear in Table 3.

TABLE 3

THE AMOUNT OF STRENGTHENING DUE TO ONE OCCURRENCE OF A CONNECTION WITH THE SPECIFIED AFTER-EFFECT; THE EXCESS WHICH THE PERCENTAGE 100S/S+D IS ABOVE 22, THAT EXPECTED FROM CHANCE PLUS NUMBER FAVORITISM

	N	0, 1	2, 3, 4	5, 6, 7, 8	F	X
One step removed	9.0	6.6	10.6	*	24.5	†
Two steps removed	7.2	5.5	4.6	*	12.7	†
Three steps removed	8.0	−0.8	6.4	*	9.3	3.7
Four steps removed	8.1	3.7	4.3	*	16.2	3.3
Five or more steps removed	5.7	3.7	7.8	12.4	12.6	2.0

* Data too scanty. For one, two, three, and four steps removed all together the figure is 4.2 but this is from only 61 cases.

† Data too scanty. Only 12 cases in the first and 15 in the second category.

Satisfying after-effects do influence the connection they follow in time even though clearly not belonging especially to them. "Here is a bonus for faithful service" has a percentage of repetition of 30.3. Its influence is small, but still +2.6 above 'no announcement.' This small influence may be low for reasons not altogether due to its lack of belonging and relevance. It was not a satisfier to all subjects because some of the subjects had special ideas concerning its import.[2]

As a check upon this evidence of the efficacy of after-effects which do not belong specifically to the connection in question, we may compute the frequencies of repetition for "You had 0 right. Very poor," "You had 1 right. Not very good," "You had 2 right. Good," etc., when in conjunction with C or F or X or NA. Table 4 presents the facts,

TABLE 4

THE INFLUENCE OF TWO AFTER-EFFECTS IN COMBINATION

0 = "You had 0 right. Very poor"
1 = "You had 1 right. Not very good"
2 = "You had 2 right. Good"
3 = "You had 3 right. Very good"
4 = "You had 4 right. Fine"
5 = "You had 5 right. Wonderful."

The entries under C, F, X, and NA are for 0, 1, 2, 3, 4, and 5 combined with C, F, X, and NA respectively.

	C			F			X			NA		
	$\frac{100S}{S+D}$	S	D	$\frac{100S}{S+D}$	S	D	$\frac{100S}{S+D}$	S	D	$\frac{100S}{S+D}$	S	D
0							23.58	50	162	27.26	290	774
1	58.96	45	30				22.02	85	301	30.88	663	1,484
2	63.50	87	50	42.86	33	44	28.96	86	211	32.95	652	1,327
3	64.50	89	48	46.88	45	51	25.72	56	104	39.63	409	623
4	68.49	50	23	47.19	43	51	37.50	15	25	38.40	154	247
5	66.67	12	6	54.55	18	15	28.57	2	5	37.65	32	63

[2] Each subject was asked at the completion of the experiment "Whenever I gave you 5 cents and said 'Here is 5 cents for faithful work,' what did it mean to you?"

10 thought that it was a reward for general good work
4 thought it meant nothing
3 thought that it was to disturb the course of the experiment
1 thought it was silly
1 thought it meant more money
1 thought it was to annoy and embarrass

using first occurrences of the specified combinations of after-effects. 2 is better than 0 or 1 in all combinations. 3, 4, and 5 are better than 2 in ten out of twelve combinations. 3, 4, and 5 are better than 0 or 1 in all combinations. C or F or X or NA with 3 or 4 or 5 is also better than C or F or X or NA alone, even when the F or X or NA is favored by the spread or scatter from a neighboring C.

As a further independent check we may inquire whether a satisfier which has ambiguous belonging displays the 'spread' or 'scatter' phenomenon. So far as our data go, this does seem to be the case. An NA just before or just after an F has a percentage of repetition of 29.6 instead of the 27.7 found for an NA remote from any satisfying after-effect.

The facts found in this experiment may be reasonably explained by the very simple hypotheses to which recent investigations of the strengthening of connections have led. These are: (1) that the mere occurrence of a connection adds a little to its strength; (2) that there exists in animals capable of learning a confirming reaction, and that this confirming reaction by whatsoever means aroused strengthens greatly the connection upon which it acts.

An occurrence plus X ('Wrong') will then produce only the amount due to occurrence alone. This after-effect never arouses the confirming reaction.

An occurrence plus 0 ("You had 0 right. Very poor") will do the same.

An occurrence plus NA (No announcement) will produce a little more strengthening, because the confirming reaction may occasionally be aroused by the subject's impression that he was right, or by accident.

An occurrence plus C ('Right') will produce much strengthening because it tends very emphatically and almost or quite universally to call forth the confirming reaction.

An occurrence plus F ("You are doing finely") will produce less strengthening than C, because it will not so often call forth the confirming reaction. Sometimes it will arouse only a sense of contentment, or pride, or zeal for the task as a whole. Perhaps also the confirming reaction called forth by

F attaches itself more to the person's general set than to the flow of word → number connections.

An occurrence plus "Here is a bonus for faithful work" will add some strength in the case of persons who are gratified by it, because there is a tendency for the confirming reaction to be put in action by any satisfier of the animal at the time. This tendency is relatively weak in sophisticated learning where the powers in control usually require relevance in the satisfier and limit the confirming reaction to the business in hand. But it is not wholly absent even there. In the learning of the lower animals it seems to be wide-spread and important.

An occurrence plus "You had — right" and any one of various commendatory adjectives with obvious dissociation of the commendation from any single connection will add some strength partly for the reason just stated in the case of "Here is a bonus for faithful work," and partly because of a more specific direction of the commendation to the flow of word → number connections.

Whether or not these explanations in accord with the simple hypothesis of slight strengthening by occurrence and much strengthening by the confirming reaction are adequate, may be left to be decided by what is now known and later discovered about the psychology and physiology of learning. The facts themselves are beyond question. Strengthening by a satisfying state of affairs is not restricted to an after-effect which is definitely attached to a certain connection by the learner's conscious awareness or by logical belonging. Within the variety used in our experiment, any satisfier tends to strengthen the connection it follows more than an annoyer does, and more than a neutral status does.

(Manuscript received June 15, 1934)

RETROACTIVE INHIBITION AND RECOGNITION MEMORY [1]

BY FRED McKINNEY

University of Missouri

Retroactive inhibition has been found in connection with a number of different kinds of tasks and materials, and several methods of learning, but up to the present time relatively little work has been done with the method of recognition in retroactive inhibition experiments. Heine [2] using the procedure of Müller and Pilzecker,[3] had a group of subjects learn nonsense syllables by the method of paired associates. These were presented by means of a memory drum. During the interpolated interval between learning and recall the subjects performed various acts depending upon the condition. During the work condition they learned other nonsense syllables; during the rest condition they either rested or studied pictures which they had to describe later. In some of her experiments in recall however, Heine had the subjects recognize the syllables they had learned instead of testing their retention by the relearning method. She used several variations of this method with the result that whenever recognition was used as a method of recall, retroactive inhibition failed to appear; whenever relearning was used to measure retention there was retroactive inhibition. In one group of experiments she had subjects read the words presented for memorizing with different accents, trochaic or iambic. One of her methods of testing retention was to record the time taken by the subject in judging whether the syllable presented was right, wrong, or doubtful. She found that the amount of retention and

[1] The first three experiments reported here were presented in a paper at the 41st annual meeting of the American Psychological Association. See *Psychol. Bull.*, 1933, **30**, p. 699.

[2] R. Heine, Ueber Wiedererkennen und Ruckwirkende Hemmung, *Zsch. f. Psychol.*, 1914, **68**, 161–236.

[3] G. E. Müller and A. Pilzecker, Experimentelle Beitrage zur Lehre vom Gedachtniss, *Zsch. f. Psychol.*, 1900, Erg. Bd., **1**, 1–300.

retroactive inhibition was proportional to the amount of time used in thinking of the accent associated with the word. All of this she regarded as evidence for the generalization that recognition memory does not depend upon association, that the recognition method of recall and the reproduction method cannot be separated on the basis of weak and strong associations.

Koffka [4] has taken this work as evidence against the theory of association in explaining all memory. He argues that already in Heine's work we have experimental evidence of a type of memory which does not involve association and implies that future evidence will further show the futility of adhering to the association concept in memory and mental life in general.

Since the time of Heine's experiment however, several other findings have suggested that memory as tested by recognition is affected by the interpolation of material between learning and recall. Strong [5] used the recognition method with advertisements. One group of subjects observed four magazines consecutively at one sitting and another examined the same material one magazine at a time, with a month intervening between presentations. A retention test four months after the first advertisements were observed showed the retention of the latter group to be superior. Lund [6] used the recognition method to test the memory of 3 series of 5-letter words. The 3 series were the same except that one letter was changed in each word in the second series and two letters were changed in each word in the third series. He found many false responses and cases of absence of recognition and explained them on the basis of retroactive inhibition due to the learning of the similar material. Gibson [7] discusses some of the implications of Heine's results for Dahl's experiments concerning the effect of sleep on retention and reports

[4] K. Koffka, *The Growth of the Mind*, Harcourt Brace, 1928, p. 382.

[5] E. K. Strong, Two factors which influence economical learning, *J. Phil. Psychol.*, etc., 1914, 11, 124–131.

[6] F. H. Lund, The criteria of confidence, *Amer. J. Psychol.*, 1926, 37, 372–381.

[7] J. J. Gibson, Retroaction and the method of recognition, *J. Gen. Psychol.*, 1934, 10, 234–236.

data on an experiment on nonsense words with 40 subjects. He found 31 percent retroactive inhibition in terms of correct responses and 53 percent in terms of errors if degree of learning which was not tested, is ignored. These amounts of retroactive inhibition are in terms of original and interpolated material which are highly similar.

It is the purpose of this study to test again the hypothesis that the recognition method does not produce retroactive inhibition, by the use of several types of materials and methods.

Method of Experiments I, II and III

Experiment I.—The traditional retroactive inhibition experimental set-up was used, the observation and recognition technique being substituted for the learning and recall technique. The set-up consisted of an experimental (work) and a control (rest) condition. In the experimental condition the subject observed a group of advertisements, the retention of which was immediately tested. Then he observed a second group, again followed by a test, and finally was retested by the same method on the first group of advertisements.

The following are the steps in the experimental condition:

1. Ten full-page advertisements [8] were taken from the *Good Housekeeping* magazine and were presented to the subject after he was instructed that he would have three seconds in which to observe each one so that he might recognize it later. He was told that the experimenter would tap with his pencil at the end of that period. Thus, each tap was a signal to turn the observed advertisement so that the next one would be available.

2. The subject spent thirty seconds observing the advertisements. Immediately after this observation these advertisements were mixed with twenty similar advertisements taken from the same magazine. The subject had to select from this group of 30 advertisements the 10 which he had just observed. He selected the advertisements by the same method that he had used in observing them, being allowed to look at each one for three seconds, and to remove from the group any that he had seen before. We shall call these 10 advertisements the original group, and it is the effect of subsequent activity upon these advertisements in which we shall be interested.

3. Subsequent to this, the subject was presented with another group of 10 advertisements, which we shall call the interpolated group, preceded by the same instructions.

4. Likewise he had to select these advertisements from a group to which 20 similar ones had been added.

5. Then, for one minute the subject counted the fives in a large number sheet consisting of the numbers five, six, seven, eight, and nine, scattered at random throughout the sheet. This activity was inserted to give the experimenter enough time to

[8] These advertisements were homogeneous as to size and color but otherwise were not particularly similar in that the cuts were of different sizes, the type, content, appeal, etc. all being different. Some of the advertisements were of nationally advertised products of great popularity; others were not.

record the number of correct choices the subject had made in his selection of the first 10 advertisements, and to mix these again for recall purposes.

6. Finally, the initial 10 advertisements were again combined with the same 20 and the subject, for the second time, had to identify them. The number of correct responses in this selection could be compared with the initial selection and the amount of forgetting thus calculated.[9]

The control condition consisted of the following similar steps:

1. The subject was given a new set of 10 advertisements which he observed for three seconds each.

2. Later, he identified them when they were combined with 20 other advertisements as he had done in the experimental condition.

3–5. During the interpolated interval of rest[10] which took the place of the second group of advertisements in the experimental condition, the subject counted the fives in a number sheet similar to that described above, for three minutes—the exact time involved in observing and identifying the second group of advertisements in the experimental condition. During this time the experimenter calculated the amount of correct recognitions and prepared the group of advertisements for the subject's recall after the rest interval.

The above-described procedure can best be understood through the following outline. OM and IM refer to original and interpolated material.

Work		Rest	
Observe OM	30 sec	Observe OM	30 sec
Recognize OM in group	90 sec	Recognize OM in group	90 sec
Observe IM	30 sec		
Recognize IM	90 sec		
Number counting	60 sec	Number counting	180 sec
Recognize OM in group	90 sec	Recognize OM in group	90 sec

Fifty subjects performed under both conditions. They were college students selected from psychology classes, many of whom were interested in advertising.

The sets of advertisements were alternated systematically in a counter balanced practice order allowing each set of advertisements to be subjected to both experimental (work) and control (rest) conditions. Also the order of presentation of the conditions was altered systematically to counter balance practice.

It is obvious that the only difference between the two conditions is the character of the material interpolated between the first and second recalls. In work a second group of advertisements was interpolated and one minute of number counting was interpolated; in rest three minutes of number counting was interpolated. The minute of number counting in work allowed the experimenter to scramble the advertisements after they had been observed, and to place them with the 20 new advertisements.

Experiment II.—It will be noticed that in Experiment I errors were ignored. It appeared at first that these errors would be negligible, but later their number seemed to be significant. This second experiment was devised to substantiate the results of Experiment I and ascertain the retroactive inhibition in terms of erroneous identifications as well as correct responses.

---

[9] It will be noted that in this fashion degree of learning is taken into consideration and not assumed to be equal in both groups.

[10] This is not a rest condition in the literal sense, but it is a condition which involves no learning and is taken up by activity found not to cause retroactive inhibition.

Further, two groups of subjects, one for the rest and one for the control condition were used in this experiment to obviate any effect which might be caused by the condition in which the same subject was used in experimental and control set-up.[11] All other details in this experiment were identical with those in Experiment I. Twenty-six different subjects were used in each of the conditions.

Experiment III.—In order to check the data collected in connection with the advertisements and to ascertain the amount of retroactive inhibition with the recognition method under very carefully controlled conditions and with more conventional laboratory material, nonsense syllables were used instead of advertisements. Whereas advertisements furnish data which are of value in demonstrating how retroactive inhibition operates with material used in our everyday learning, they are not the best material in an analytic study where as many factors as possible are to be controlled. In the case of nonsense syllables, the subject has but one stimulus to which to respond and learn and his set can be easily directed, whereas in the advertisement there is an innumerable quantity of them and it is hard to know which he is really using in identifying the recognized material. In view of these considerations, it was thought well to check the above results with this more artificial material in ascertaining the amount of retroactive inhibition occurring among impressions gained not through many trials as memories usually acquired in the laboratory, but through single impressions. It must be remembered that learning of the everyday sort, as many writers have pointed out, is that which occurs as a result of single trials and casual impressions.

This experiment also allows a comparison of the phenomenon being tested in the case of both meaningful, well integrated, everyday material with relatively non-meaningful, artificial laboratory material. The methods used in the case of the two materials are identical, making comparison valid.

Three sets of 10 groups of nonsense syllables with 20 percent association value, as found by Glaze,[12] were neatly typed on three-by-five cards. Sixty other nonsense syllables, all typed for mixing purposes, 20 for each of the above groups of 10, were prepared.

The same experimental set-up was used here as in Experiment I. One pack of nonsense syllables was used for the original observation, then 20 nonsense syllables were used for its mixing group. Another pack of 10 was used for interpolation and another group of 20 for mixing. Then the third group was used in rest condition. The groups were alternated systematically so that each set served 12 times in each position and the results were not a function of the lists.

The subjects were of the same type as mentioned above. As in Experiment I, the rest and work conditions were counter balanced systematically. The same exposure time was used and the subjects turned the cards as they did the advertisements. Errors and correct responses were recorded for both the first and second recognitions of the original and interpolated groups of the work and rest conditions. The amounts of forgetting and of retroactive inhibition were calculated.

---

[11] Since this was the case, it was necessary to have only two groups of advertisements instead of three—one for the original and one for the interpolated group. The original set of advertisements in this case was more homogeneous; for example, two advertisements of one company were included. This was done purposely to ascertain, in a somewhat general fashion, whether similarity is effective in these conditions as in the more artificial conditions previously used in retroactive inhibition experiments.

[12] J. A. Glaze, The association value of nonsense syllables, *Ped. Sem. and J. Genet. Psychol.*, 1928, **35**, 255–267.

Results of Experiments I, II and III

Table I shows the mean number of advertisements correctly recognized immediately after learning and after an

TABLE I

Means and Sigmas of Correct Recognitions of Advertisements in Experiment I

	Correct Responses			
	Rest		Work	
	Mean	σ	Mean	σ
Recall 1	7.84	1.36	7.30	1.92
Recall 2	6.58	1.65	5.92	1.57
Percent Recalled	83.92		81.09	
Retroactive Inhibition	03.37%			

interval for the work and rest conditions in Experiment I. These two recalls are designated as Recall 1 and Recall 2, respectively. The percentage of recall, using Recall 1 as a standard and the percentage of relative retroactive inhibition are also included. Sigmas of the distributions for Recalls 1 and 2 under rest and work are included.

It will be seen that there is a rather large percentage of delayed recall and a small amount of retroactive inhibition. The immediate and delayed recall under the work and rest condition are not significantly different from the statistical point of view but the interpolated learning does decrease the amount recalled to a small degree.

TABLE II

Means and Sigmas of Correct Recognitions and Errors in Advertisements. Experiment II

	Correct Responses			
	Rest		Work	
	Mean	σ	Mean	σ
Recall 1	9.16	1.09	8.76	1.24
Recall 2	9.16	.948	8.16	1.19
Interpolated Recall			8.57	1.33
Percent Recalled	100		93.15	
Retroactive Inhibition	6.85%			

	Errors			
	Rest		Work	
	Mean	σ	Mean	σ
Recall 1	.615	.594	.384	.642
Recall 2	.731	.754	.923	.904
Interpolated Recall			.19	.574
Relative Increase in Recall 2	18.86		140.36	

Table II, giving data from Experiment II, shows the same sort of results,—a smaller percentage of delayed recall under the work condition than under the rest condition, and a slight amount of retroactive inhibition.

The differences between the recall of the rest and work conditions are not significant from the statistical viewpoint, but they show the same tendencies as exhibited by Table I. Additional evidence showing the same trend is shown by the calculation of the magnitude of errors occurring in recall. An error was recorded whenever the subject selected the wrong advertisement in the recognition recall test. There is a rather small increase in errors from Recall 1 to Recall 2 during the rest condition, and rather large increase during the work condition. In fact, the increase of errors is doubled due to the interpolation of additional advertisements.[13]

It may be concluded from these two experiments that the interpolation of similar material between the learning of a group of advertisements by observation and the recognition of those advertisements diminishes the percentage of recognition. In other words, interpolation of homogeneous material causes a small amount of retroactive inhibition in the memory of advertisements using the recognition method of recall.

Table III contains similar data on the immediate and delayed retention of nonsense syllables, the percentage of recall, and the amount of retroactive inhibition. The same type of results is found in connection with the nonsense syllables, namely a small amount of relative retroactive inhibition, but a greater amount than that found with the advertise-

[13] There are two possible methods of calculating retroactive inhibition in this case, namely in terms of the percentage of increase of errors in Recall 2 over Recall 1 or in terms of the percentage Recall 2 is of Recall 1. The former yields 18.86 percent increase in errors under rest condition and 140.36 percent increase in errors under work condition. This yields a very large difference and a relative retroactive inhibition of 644 percent—a figure which is not comparable with the usual mathematical expressions of retroactive inhibition. The latter yields 118.86 percent, the percentage Recall 2 is of Recall 1 for rest condition and 240.36 percent for work condition resulting in 102.22 percent of retroactive inhibition. Because of the difficulty of obtaining retroactive inhibition scores in terms of these errors which compare with those obtained in terms of correct responses they have been omitted from the table and the reader is allowed to see from inspection the effect of the interpolated advertisements.

TABLE III

MEANS AND SIGMAS OF CORRECT RECOGNITIONS AND ERRORS OF NONSENSE SYLLABLES. EXPERIMENT III

	Correct Recognitions					
	Rest			Work		
	Mean	σ		Mean		σ
Recall 1	6.50	1.54		6.78		1.77
Recall 2	5.89	1.48		5.25		1.49
Interpolated Recall				6.39		1.57
Percent Recalled	90.61				77.43	
Retroactive Inhibition			14.54			

	Errors					
	Rest			Work		
	Mean	σ		Mean		σ
Recall 1	4.56	2.14		4.19		1.88
Recall 2	6.11	2.46		6.81		1.91
Interpolated Recall				4.83		1.95
Relative Increase in Recall 2	(1) 33.99				62.52	
	(2) 133.99				162.52	
Retroactive Inhibition			(1) 83.93			
			(2) 21.29			

ments, in terms of correct responses. In terms of errors [14] however, the nonsense syllables show a smaller amount of inhibition than the words. This can probably be explained in terms of the material used. There is a greater amount of variability in connection with errors made in selecting advertisements than in connection with the nonsense material. Therefore, greater emphasis probably should be placed on the correct responses in comparing the two types of materials.

EXPERIMENT IV

After these three experiments had been completed and the data had been compiled, it was decided to carry out one final experiment duplicating as nearly as possible the methods of Heine, particularly the degree of learning. Since in these three experiments just described the material was observed for a short interval and the subject did no formal repeating it might be argued that the results differed from those of Heine's because of a difference in the degree of learning in the two

[14] It will be noted that here as in Table II the relative increase in Recall 2 is calculated by two methods namely, in terms of the percentage of the increase of errors in Recall 2 over Recall 1, and in terms of the percentage Recall 2 is of Recall 1. Likewise, there are two calculations for the amount of retroactive inhibition.

cases. It might be thought that under such conditions in which learning is slight, forgetting is easy. Since however, it is the purpose of this experiment to ascertain if retroactive inhibition is present with the recognition method of testing memory, all types of memory—strong and weak—which may be tested by recognition should be studied. It seems logical that the least favorable condition for the appearance of the phenomenon of retroactive inhibition in connection with the recognition method is the case in which the material is well fixated to a criterion and later merely selected from a larger group of material. If the phenomenon occurs then when the learning is extensive and the material is well fixated and when the requirements for testing the retention of this learned material is easily met, then retroactive inhibition will probably be found under all conditions.

This experiment meets these conditions in that the material was well learned and retention was tested by recognition.

Method.—The same type of material as that used by Heine was employed namely, nonsense syllables. The nonsense syllables were selected from Glaze's tables and had 53.33 percent association value. This degree of association value was chosen because it represented a mid-point in association value of all possible nonsense syllables and thus the syllables would have a greater likelihood of resembling Heine's.

The method of presentation was as similar to Heine's as possible considering improvements in memory experimentation technique since the time of her experiment. A McGeoch electric memory drum served as the exposure apparatus. The constant exposure time for each word was two seconds. This differs from Heine's technique in that she allowed the subject to learn at his own speed and thus the time of exposure was suited to the subject. This would allow variations in degree of learning which are undesirable so a departure in the direction of control of exposure time was made in the case of this experiment.

The instructions given to the subjects were similar in this experiment to Heine's. They were told not to rehearse the lists between presentations and to respond only to the syllable appearing at the time of exposure other than to past syllables. The method was the anticipation method rather than the method of allowing the subject to read the syllables until he thought he had acquired them. Whether the subject had learned the syllables or not was decided objectively in terms of his ability to anticipate the whole list one time perfectly. These are merely the common present-day methods of controlling memory experiments and a more rigorous procedure than Heine's. A greater degree of learning should result and this condition should produce a more exacting test of the existence of retroactive inhibition.

The interpolated activity under the work condition consisted of another list of similar nonsense syllables; under the rest condition, the material was the same as Heine's namely, pictures to be observed for an interval and described. The pictures were colored covers from *The Literary Digest* and consisted of landscapes, still life

pictures, and social scenes. The interpolated interval was a constant of 9 minutes similar to Heine's 8, 9, or 9½ minute intervals.

In testing retention, the subject was told that the syllables learned in the first list would appear again among others and he must distinguish between the new and the old. He was, as in the case of Heine's experiment, allowed to mention those concerning which he was doubtful. The 16 old syllables were scattered at random among 16 new ones and this list was presented to the subject four times to get a reliable measure of recognition. The subject was supposed to observe each syllable and react to it as soon as possible with the words "old," "new," or "doubtful." No record was taken of time of response. After the subject had gone through the recognition test four times he relearned the first list to a criterion of three perfect trials so that a comparison might be made of the recognition and relearning methods of measuring retention in the computation of retroactive inhibition.

The general set-up of the experiment was strictly comparable to the three described above with the specific differences mentioned. The rest condition consisted of learning the list by anticipation—observing and describing pictures for 9 minutes, recognizing and relearning the original list. The work condition consisted of the same with the exception that the observation and description of pictures were substituted by the learning for 9 minutes of another list of nonsense syllables and a recognition of this list just prior to the conclusion of the experimental period. It was as similar to Heine's as was possible and still allowed for all controls which are necessary to obtain reliable accurate quantitative data.

Different subjects were used for the two experiments. They were undergraduates and all but a few had never been in a memory experiment before. They were naive as to the purpose of the experiment. Twenty-two were used in the rest, and twenty-one in the work condition.

TABLE IV

RECOGNITION SCORES AND PERCENTAGES OF RETROACTIVE INHIBITION FOR EXPERIMENT IV

	Rest	Work	% R.I.
4 Errors in Recognition	.681± .29	.809± .24	18.79
3 Errors in Recognition	.860± .26	.905± .19	05.23
2 Errors in Recognition	.860± .28	1.330± .24	54.65
1 Error in Recognition	1.727± .30	1.857± .41	07.53
Total Recognition Errors	8.770±2.31	10.476±1.41	19.45

Results.—Table IV shows the scores for all four recognition tests, the total recognition errors for both groups and the relative percentage of retroactive inhibition for all of these measures. The recognition scores show the total number of errors on the four recognition tests. The cases in which the subject made a consistent error on all four tests, that is, either said a syllable belonged to the previous list when it did not or vice versa, were recorded in the table as four errors in recognition. When the subject made an error in recognizing a syllable three times out of the four trials it was recorded as

three errors in recognition and likewise for two errors and one error in recognition. There were so few cases where the subject used the response 'doubtful' that these were not recorded in the table. In calculating the number of erroneous responses in the four recognition tests the doubtful responses were not included. The listing of total errors in the table is a total of all the errors made on all four tests.

It can be seen from the table that the percentage of relative retroactive inhibition in terms of the total recognition errors is 19.38. It is about the same for the cases where the subject made four errors on four successive tests of his ability to recognize the word. The amount of retroactive inhibition in terms of the cases where the subject made one error or three errors in four is lower. In the cases in which he made two errors in four the amount of retroactive inhibition is much higher. This can probably be explained in terms of the assurance the subject had concerning the recognition of the syllable. In cases in which he was not sure whether he had seen the syllable or not he would say that he had 50 percent of the time and that he had not 50 percent of the time, and thus he would make two errors in the four tests. It might be expected that an interpolated list would cause the subject to be more doubtful of some syllables and therefore it is not surprising that the two correct responses out of four category in our data shows a higher retroactive inhibition than those cases in which the subject is more positive of his reaction and gives three correct or one correct response.

It was intended at first to compare the recognition with the recall and relearning methods of testing retention using the same material and subjects but it is apparent that the four recognition tests complicated the results in both conditions by acting as a review, so that the retroactive inhibition in terms of recall and relearning is much less than one would expect with this type of material [15] and cannot validly be compared with the recognition scores.[16]

[15] J. A. McGeoch, The influence of four different interpolated activities upon retention, this JOURNAL, 1932, 15, 662–680.

[16] Under the rest condition the subjects required an average of 25.73 ± 1.97 trials to learn; under the work condition 20.86 ± 1.35 trials. This difference is not

Discussion

In the light of these four experiments there seems little doubt but that retroactive inhibition occurs with these materials when learning takes place through the methods used and memory is tested by the recognition method. Retroactive inhibition occurred to some degree in each case by each measure used. The amounts of retroactive inhibition produced by these methods appear to be smaller than those usually found.

It is clear that these results conform well with other data. The smaller amounts of retroactive inhibition with the advertisements are what might be expected in terms of the other results on retroactive inhibition in using meaningful material.[17] Greater amounts of retroactive inhibition with nonsense syllables as shown by correct score data are in accordance with previous findings. Greater retroactive inhibition with more similar material is also in accordance with previous findings.

As to the theoretical significance of these data it is evident that these data agree with the transfer theory in that there was more retroactive inhibition when more homogeneous advertisements were used.

Since the relationship of this experiment to association theory has been raised by Heine and later by Koffka, it might be briefly considered here. There is here no evidence to show that retroactive inhibition varies greatly in nature when recognition is used as the method of measurement rather than recall. There is a smaller amount of inhibition with the former

significant and shows if anything a slight tendency for the subjects under the work condition to be better learners. Under the rest condition the average trials to relearn were 4.32 ± .55 and under the work condition 3.80 ± .63. In terms of these measures there was 12.04 percent relative retroactive inhibition. The group saving score for the rest condition was 83.21, for the work condition 81.78. The percentage of retroactive inhibition in terms of this measure was 1.72. The individual saving score under the rest condition was 80.41 ± 3.30; under the work condition 79.00 ± 2.71, with a relative amount of retroactive inhibition of .017. The recall score under the rest condition was 12.09 ± .75, under the work condition 10.66 ± .96, and the relative retroactive inhibition in terms of these measures was 11.83.

[17] J. A. McGeoch and Fred McKinney, Retroactive inhibition in the learning of poetry, *Amer. J. Psychol.*, 1934, **46**, 19–33.

method but these data offer no basis for the two methods of measuring memory as being fundamentally distinct. There is no doubt that the methods differ [18] but more evidence is necessary if it is to be argued that this difference is fundamental. The various theories of recognition memory have been discussed fully elsewhere,[19] and all that can be added here as the result of this experiment is the statement that these data offer no cogent evidence for a qualitative separation of this type of memory from other types.

These data go further in showing that the phenomenon of retroactive inhibition is operative not only in the learning of skills and complex tasks and materials in and out of the laboratory, but also in memories which are the result of single trials and observations of the same order as everyday memories. Probably most of our mental content is gained incidentally by impressions which are not the result of repetition but of single events and it is always important to know whether the principles found operative in the more formal kinds of learning are also operative in these daily situations. These data imply that the forgetting of names, faces, places, objects, acquired incidentally in the daily course of events is due to some extent at least to interpolated events.

Summary

The typical retroactive inhibition set-up was used in these experiments. In the first experiment fifty subjects observed 10 advertisements for 3 seconds each. Then, these 10 advertisements had to be identified among 20 others. In the work condition this was followed by the same procedure on another set of advertisements and in the rest condition the subjects counted numbers on a number sheet for the same length of time. In both conditions this interpolation was followed by the duplication of the original recall and the retention compared. The same subjects were used in the

[18] Woodworth discusses some interesting differences between the two methods. See R. S. Woodworth, *Psychology*, Revised, 1929, New York, Henry Holt and Company, 108–109.

[19] See references 1068–1095 in J. A. McGeoch, The psychology of human learning: A bibliography, *Psychol. Bull.*, 1933, **30**, 1–62.

rest and work conditions, but different sets of advertisements were used and systematically counter balanced to take care of practice and other effects. Forgetting under rest and work conditions and retroactive inhibition were calculated.

In the second experiment the same set-up was used on different advertisements with the exception that different subjects were used under rest and work. In this way the original results were checked.

In the third experiment nonsense syllables neatly typed on cards three by five inches were substituted for the advertisements. The same procedure and calculations were made.

A fourth experiment duplicated Heine's result with slight improvement on control namely, having the subject learn the original lists to perfection before interpolation.

The results show small but consistent amounts of retroactive inhibition with all material used. The similar advertisements produced the greatest amount of retroactive inhibition in terms of number of advertisements correctly selected, the dissimilar advertisements the least. Nonsense syllables show a slightly greater amount than the advertisements.

These results have theoretical significance in that Koffka used Heine's previous negative results on this same problem as evidence for a gestalt rather than an associational explanation of recognition. These data in contradicting Heine's findings vitiate Koffka's argument.

(Manuscript received June 19, 1934)

BACKWARD CONDITIONED EYELID REACTIONS

BY HULSEY CASON

University of Wisconsin

As a minor part of a study of conditioned eyelid reactions in 1922 (**5**; **6**), the writer attempted with one subject to establish backward conditioned eyelid reactions (**5**, 191). We found no evidence in favor of backward conditioning, and several other lines of evidence (**3**; **4**; **7**; **9**; **10**; **15**; **13**; **22**; **2**; and **11**) seem to indicate that backward conditioning in a strict sense does not occur. Several writers (**16**, 393–394; **1**; **17**, 381–384; **20**; **23**; **24**; and **14**), however, are in one way or another favorable to the concept of backward conditioning, and the present experiment was designed to furnish additional data on this matter and on the more general problem of remote association. The questions at issue have been admirably set forth in recent articles by Guthrie (**13**) and Wenrick (**22**).

Apparatus and Procedure Used in the Present Experiment.—A spectacles apparatus (**5**, 162–163) similar to the apparatus used by a number of investigators was adjusted to S's head, and the end of a light lever was attached by a piece of sticking plaster to S's left eyelid just above the free margin. With each natural eyelid reaction, a piece of platinum wire on the far end of the lever came in contact with a second piece of platinum wire placed in a favorable position. During the conditioned response (C–R) training period this contact apparatus was used to close an electric circuit. The contact between the two platinum wires occurred with each eyelid reaction, and the circuit was closed when the eyelid had moved approximately four-fifths of the total distance. The advantages of this type of eyelid apparatus have been recognized by a number of investigators.

S and E (the writer) were in different rooms, and the rooms were connected by a cable of wires. S was seated during practically all of the experiment, and was reading or studying during practically all of the C–R training period. S could signal E to come into his room whenever he wished.

The C–R training was carried out in a simple manner. With each natural eyelid reaction, the spectacles-lever-contact apparatus closed an electric circuit, and this produced a click sound in a telegraph relay in S's room and 4 ft in front of him. The click sound was fairly loud for a telegraph relay, but not loud enough to produce a wink in any S at the beginning of the experiment. The total number of winks at different times during the experiment was automatically registered in E's room by an electric counter, the counter being in circuit with the make side of the relay in S's room.

Each repetition of the C–R training stimuli included the following four events, which were not analyzed by any of the S's. (1) A natural intra-organic stimulus leading to a natural eyelid reaction. (2) The natural eyelid reaction itself. (3) The auditory click sound stimulus made by the relay in S's room. (4) The auditory response in S's body. With each repetition of these C–R training stimuli, there were two stimuli and two responses, and we have previously discussed (8) several possibilities of different connections between stimuli and responses in situations of this kind.

In each repetition of the C–R training stimuli, the conditioned sound stimulus was given when the original wink response was approximately four-fifths completed. As far as the central nervous system was concerned, the conditioned sound stimulus in each repetition of the C–R training stimuli was given immediately after the original winking reaction.

The connections and wiring were such that by means of a four-pole-four-throw switch, I could (1) break the connection between S's winking-contact apparatus and the sound producing relay in his room, and between S's relay and my electric counter, (2) place the switch in a neutral position, or (3) without delay simultaneously produce exactly the same sound stimulus in S's room and start a Dunlap chronoscope in my room. The pointer of the chronoscope was stopped and clamped if and when the S winked. The chronoscope showed whether S winked, and if he winked the time which elapsed between the sound stimulus and the moment when his eyelid was four-fifths closed. The switch could be moved back to the first position and the self-training could be continued as before.

The S's were 8 junior and senior university students, approximately 20 years of age, 7 women and 1 man. All had taken at least 2 courses in psychology, but verbal reports at the end of the experiment showed that all except No. 5 were naive as far as the present experiment was concerned.

The routine of the experiment with most of the S's was as follows.

1. S was seated in a comfortable Morris chair, where he generally remained throughout the experimental period except when he occasionally stood up to avoid fatigue from prolonged sitting. (The experiment continued while he was standing up.) The spectacles apparatus was put on and adjusted. S was instructed to make no attempt to wink or not to wink. No suggestions were given on the nature of the problem being studied. The experiment was carried out during the period of final examinations in June, and all S's gave their best attention to reading and studying during the experiment.

2. S winked naturally for a period of 2–4 hrs, and as described above gave himself an auditory stimulus every time he winked. I recorded the total number of eyelid reactions which had occurred 15 or 20 times during the experiment.

3. During the latter part of the experimental period, the same sound stimulus was given alone 35–40 times after irregular intervals, at an average rate of approximately 2 trials a min. A record was made of the time of each eyelid response that occurred, if the response occurred within a period of 1 sec. The C–R training was continued in the intervals between the C–R trials.

4. Several measures were made of S's fastest voluntary winking time (V. W. T.) to the same sound stimulus. No ready signal was given in these V. W. T. trials. While these trials were being made, the sound was not produced by S's winking.

All time measures in (3) and (4) above were for the interval between the click sound from the relay and the moment when the eyelid had moved approximately four-fifths of the total distance.

5. During the latter part of the period of C–R trials, I asked S whether the apparatus was working all right, and whether the sound had occurred at any time without his winking. At the end of the experiment, S was carefully questioned to determine the kind of verbal processes which had been present in him during the experiment.

Summary of Individual Results.—Subject No. 1, *Miss J. M.*, 7:15 P.M. to 12:05 A.M. Between 7:45 P.M. and 11:39 P.M., 465 natural eyelid reactions occurred, and each wink was followed by the conditioned sound stimulus. Between 10:18 P.M. and 11:39 P.M., the sound was given alone 31 times, and no C–R occurred within one sec on any trial. Average natural winking rate during C–R trials was 1 wink every 36 sec. At 11:20 P.M., I asked S if the sound had occurred without her winking at any time, and she said she had not noticed it. The sound had occurred without her winking 17 times. S paid attention to the sound during the whole experiment, however, and on 2 or 3 occasions, when she heard the sound near the end of the period during which the C–R trials were made, S thought that she must have just winked. There was apparently no impulse to wink for the sound alone. Questioning at the end of the experiment showed that S has no insight into its purpose and nature.

S No. 2, Mr. R. H., 1:15 P.M. to 6:20 P.M. Between 1:45 P.M. and 5:51 P.M., 3402 natural winks, each followed by the conditioned click sound. Between 5:27 P.M. and 5:49 P.M., sound given alone 38 times, and no C–R within 1 sec on 35 trials. Three winks to sound alone occurred on the 3d, 29th, and 36th trials, the times being 286, 192, and 258 σ. (σ = .001 sec.) Average natural winking rate during C–R trials, 1 wink every 10 sec. On some occasions the conditioned sound stimulus may have been given when a wink had already been, or would have been, initiated. At end of experiment S said he did not think the sound had occurred alone without his winking. No insight into experiment.

S No. 3, Miss R. H., 7:30 P.M. to 12:15 A.M. Between 8:26 P.M. and 11:40 P.M., approximately 1739 natural winks, each followed by sound. Between 11:26 P.M. and 11:42 P.M., 37 C–R trials, and no response within 1 sec. on 35 trials. Two eyelid responses occurred on trials 13 and 35, times 771 and 631 σ. Average natural winking rate during C–R trials,

1 wink every 13 sec. In reply to questioning at 11:32 P.M., S said she had not noticed that sound had occurred alone without her winking. Sound had occurred alone 21 times. However, on some occasions she thought there had been too long an interval between her previous wink and the sound. Between 11:32 P.M. and 11:42 P.M., sound given alone 16 times, and S now noticed that it sometimes occurred without her winking. She said later that during the latter part of this period, during which the C–R trials were made, the sound I made without her winking was weaker than the sound she made by winking. S had no impulse to wink in response to the sound. At end of experiment, 15 measures made of V. W. T. without relay signal, eyelid responses occurred on all trials, Range 161–476 σ, Median 228 σ. No insight into experiment.

S No. 4, Miss R. R., 8:40 A.M. to 12:35 P.M. Between 9:18 A.M. and 12:04 P.M., 572 natural winks, each followed by sound. Between 11:39 A.M. and 12:04 P.M., sound given alone 40 times, and no C–R on 35 trials. Five winks occurred on trials 4, 5, 8, 20, and 22; times 898, 472, 626, 473, and 126 σ. Average natural winking rate during C–R trials, 1 wink every 29 sec. At 11:53 A.M., I asked S if sound had occurred without her winking, and she said she thought it had not. Sound alone had been given 29 times after irregular intervals during the preceding 14 min. Between 11:53 A.M. and 12:04 P.M., sound given alone 11 times, and S noticed that the sound occurred without her winking on some of these occasions. She said the sound I made without her winking seemed weaker than the sound she made by winking, although the physical sound was exactly the same on the two occasions. At end of experiment, 10 trials made of V. W. T., responses occurred on all trials, Range 162–898 σ, Med. 262 σ. No insight into experiment.

S. No. 5, Miss J. F., 1:40 P.M. to 5:02 P.M. Between 2:16 P.M. and 4:27 P.M., 2423 natural winks. Between 4:09 P.M. and 4:28 P.M., sound given alone 39 times, and no C–R on 28 trials. Eleven winks occurred on trials 3, 4, 6, 7, 8, 12, 23, 24, 26, 28, and 38; the times being 341, 719, 535, 551,

556, 457, 111, 860, 548, 872, and 720 σ. Average natural winking rate during C–R trials, 1 wink every 3 sec. It is quite probable that some of these eyelid reactions had already been initiated when the conditioned sound stimulus was given alone. At 4:22 P.M., S had not noticed that sound had been given alone, although it had occurred alone 26 times during the preceding 13 min. Between 4:22 P.M. and 4:28 P.M., sound was given alone 13 times, and S later said that during this time she did notice that the sound sometimes occurred without her winking. At end of experiment, 11 measures made of V. W. T., responses occurred on all trials, Range 186–292 σ, Med. 232 σ. This S was not entirely naive. She thought the experiment was on the conditioned reflex, although she did not think about the topic of backward conditioning.

S No. 6, Miss M. H., 1:30 P.M. to 6:25 P.M. Between 2:03 P.M. and 6:01 P.M., 842 natural winks, each followed by sound. Between 5:37 P.M. and 5:59 P.M., sound given alone 40 times, and no C–R on 39 trials. One wink occurred on 32nd trial, time 425 σ. Average natural winking rate during C–R trials, 1 wink every 24 sec. At 5:52 P.M., after sound had been given alone 31 times, S said she thought sound had not occurred alone without her winking. At end of experiment, 11 measures made of V. W. T., responses occurred on all trials, Range 126–429 σ, Med. 176 σ. No insight into experiment.

S No. 7, Miss J. L., 7:25 P.M. to 12:30 A.M. Between 7:54 P.M. and 12:03 A.M., 7264 natural winks, each followed by sound. Between 11:47 P.M. and 12:03 A.M., sound given alone 39 times, and no wink within 1 sec on 29 trials. Ten winks occurred on trials, 5, 6, 9, 10, 15, 17, 20, 27, 36, and 37; the times being 189, 364, 36, 129, 83, 25, 15, 295, 47, and 114 σ. Several of these eyelids reactions were obviously not due to the conditioned sound stimulus. Average natural winking rate during C–R trials, 1 wink every 1.1 sec. At 11:55 P.M. and again at 12:03 A.M., I asked S if sound had occurred alone without her winking, and on both occasions she said it had not. Sound had occurred alone 24 times on the first and 39 times

on the second occasion. At end of experiment, 17 measures made of V. W. T., responses occurred on all trials, Range 147–325 σ, Med. 187 σ. Two measures of 80 and 82 σ were not used in above calculation, but including them would change Med. to 178 σ. No insight into experiment.

S No. 8, *Miss G. H.*, 8:40 A.M. to 12:50 P.M. Between 9:31 A.M. and 12:20 P.M., 1020 natural winks, each followed by sound. Between 11:58 A.M. and 12:20 P.M., sound given alone 40 times, and no C–R on any trial. Average natural winking rate during C–R trials, 1 wink every 29 sec. At 12:11 P.M. and again at 12:20 P.M., I asked S whether sound had occurred alone without her winking, and she said she thought it had not. Sound alone had been given 24 times on the first and 40 times on the second occasion. At end of experiment, 14 measures made of V. W. T., responses occurred on all trials, Range 231–877 σ, Med. 503 σ. No insight into experiment.

Aside from the bearing on the present C–R experiment, the natural rate of winking of our S's is of some interest in itself; and in Fig. 1 we have plotted the total number of natural eyelid reactions at different times during the experiment for 7 of the 8 S's. We do not have these exact figures for the 3d S because the electric counter 'refused' at the beginning of the experiment with this S. The curves show the total number of winks which occurred before and after the C–R tests were begun. There were large differences in the average winking rates of the 7 S's, and some differences in the winking rate of the same S at different times during the experiment (Cf. **19**; **18**).

Summary and Conclusions.—S winked naturally for a period of 2–4 hrs, and a sound stimulus was automatically given immediately after each natural eyelid reaction occurred. For the 8 S's, there was a total of 17,727 repetitions of these C–R training stimuli.

In the latter part of the C–R training period, my giving the same sound alone did not call out eyelid reactions in any regular fashion, and on these C–R trials no wink occurred within 1 sec in 2 out of the 8 S's. In 304 C–R trials, the

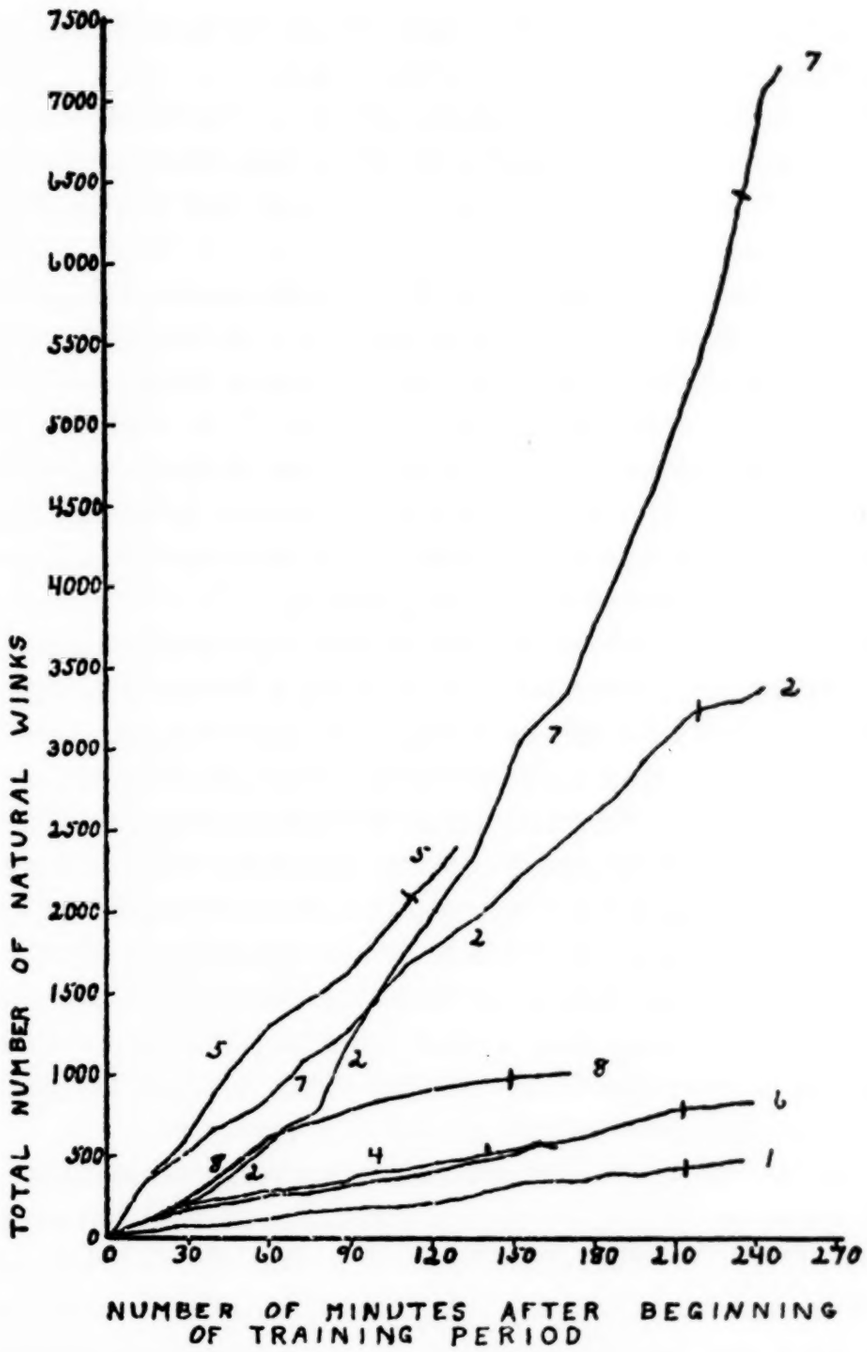


FIG. 1. Total number of natural eyelid reactions at different times during the experiment. The number by each curve refers to the number of the S. The cross mark on each curve indicates the time the C–R trials were begun.

8 S's gave a total of 32 winks (11 percent), and 21 of these 32 winks occurred in the 2 S's, Miss J. F. and Miss J. L., who had the fastest natural winking rates.

At the end of the experiment, 6 S's were asked to wink as

quickly as they could when they heard the sound, and eyelid responses occurred on all of these trials.

When the sound was given alone in the C–R trials, the S's generally had no impulse to wink, and on most occasions they did not even know that the sound had occurred alone without their winking.

An earlier experiment on the conditioned eyelid reaction (**5**) showed that the eyelid response could be conditioned to an auditory stimulus, when the two stimulus-response patterns overlap each other temporally in the C–R training. The present experiment shows that a connection between an auditory stimulus and an eyelid reaction is generally not formed when the auditory stimulus comes immediately after the eyelid reaction in the C–R training.

However, after the auditory stimulus-and-response had occurred immediately after the winking stimulus-and-response a large number of times in the present experiment, some kind of conditioning may have occurred, because the presence or absence of the eyelid stimulus-and-response made a difference in the apparent intensity of the auditory stimulus-and-response. The sound occurring alone seemed weaker than the sound immediately preceded by an eyelid reaction. The auditory response was apparently conditioned to the organic pattern of the preceding eyelid reaction; the eyelid reaction seems to have acquired the ability of increasing the intensity of the immediately following auditory response.

Another possible explanation, however, is that before the C–R training began, a reflex connection existed between the facial pattern of the eyelid reaction and the tension of the muscles in the middle ears. If this explanation be the correct one, then one would probably conclude that no noticeable conditioning of any kind occurred in our experiment. I did not check the possibility of this reflex connection between the eyelid and auditory responses existing before the beginning of the C–R training period.

In the writer's earlier experiment on the conditioned eyelid reaction (**5**), an electric shock applied just below the eye called out original eyelid reactions, and in the C–R training a sound

stimulus was given with this shock-stimulus-and-eyelid-reaction. The intensity of the shock was the same throughout the experiment. At the beginning of the C–R training period, the shock alone, in the case of 4 S's, seemed of the same intensity as the shock in the shock-sound pattern; but after the shock and sound had occurred together in the C–R training a large number of times, the shock alone seemed weaker than the shock in the shock-sound pattern (**5**, 192–193). Bernstein (**2**) has made similar though not identical observations.

Discussion of Results on Backward Conditioning.—Switzer (**20**) and Wolfle (**24**) have suggested that one of the reasons I did not obtain any backward conditioning in 1922 (**5**, 191) was that in the C–R training the conditioned sound stimulus was given within the refractory period of the eyelid reflex. If this were a valid explanation, there should be some kind of tendency for ease of conditioning to follow the general curves of the absolute and relative refractory period. But the results on backward conditioning do not show that this is the case. The processes involved in these experiments are also not limited to the reflex level. Switzer and Wolfle seem to have made their suggestion in regard to the refractory periods before they knew what the refractory periods of the eyelid reflex were (See **12**; **21**).

Switzer (**20**) claimed backward conditioning in 18 out of 20 S's using the eyelid reaction, in 2 out of 5 S's with the patellar response, but in none out of 5 S's with the galvanic skin response. He does not give his complete results, but states that the criterion of conditioning used was 3 or more successive C–R's. If such a criterion had been used in the present experiment, we would have concluded that one of our S's, No. 5, was conditioned, although no C–R's occurred within 1 sec in 28 out of 39 trials.

Switzer's positive results may have been due to the verbal sets (**11**) of his S's, and it seems probable that little conditioning would have occurred if the verbal processes of his S's had been otherwise engaged. Complications are liable to result when the verbal processes are neglected altogether.

In our present experiment, spending just a few seconds in telling the S's to wink as quickly as possible when they heard the sound produced eyelid reactions with a reaction time of less than 1 sec on all trials. Seventy eight trials of V. W. T. were made with the last 6 S's.

Wolfle's experiment (**24**) has also been interpreted as favoring the existence of backward conditioning. She attempted to condition a movement of one of the middle fingers to a sound by applying an electric shock to the finger and hand and giving the conditioned sound stimulus before, simultaneously with, or after the electric-shock-and-succeeding-withdrawal-movement. Each S was given 2 or more 40-min practice periods on different days. Out of the 72 presumably intelligent college students who served as S's, 25 were not able, even after considerable training, to learn to move their finger when they heard a sound. With 20 of the 47 remaining S's, the more they were 'trained' the worse they got. With the 72 S's, 10 C–R trials were made during the last 20 min of each practice period. When the conditioning shock stimulus was given 1 sec before the conditioned sound stimulus in the C–R training, the percentages of C–R's obtained in successive practice periods were 0–0–0–0 for 1 S, 70–30–0–0–0 for a 2d S, and 30–10–0–30 for a 3rd S. When the sound was given .3 sec before the shock in the C–R training, the percentages of C–R's obtained in successive practice periods were 100–100–100–100–10–50–0–10–0 for 1 S, 0–0–0–0–0 for a 2d S, and 100–100–100–100–100 for a 3rd S. In other procedures, some other S's gave percentages of 20–70–0–80, 0–0, and 0–10–50–80. In a control group of 6 additional S's, no C–R tests were made until the 4th and 5th practice periods; and in the 5th practice period, out of 6 S's, 3 gave 100 percent of C–R's, and 2 gave 5 percent of C–R's. It seems that Wolfle's procedure in averaging percentages of 100 and 5 has led to an erroneous impression.

One can scarcely explain the extreme variability in some of Wolfle's results by physical and physiological factors in the experiment, such as, for example, the variation in resistance at the electrodes. One is almost forced to conclude

that several of the more fundamental causal factors in this experiment were not controlled and may not have been suspected. It is possible that some of her S's might have moved their finger when the sound occurred before the C–R training began. General principles of conditioning and learning cannot be established by experiments in which observations are strictly limited to the behavior of, for example, the hand, when the verbal activities which afford a fairly satisfactory explanation of what the hand does are located elsewhere in the body (**11**).

Bernstein (**2**) has carried out a carefully controlled experiment on conditioned eyelid reactions, but he seems to have used too few repetitions of the C–R training stimuli to obtain conditioning in many cases. An electric shock called out original eyelid responses, and in the C–R training the conditioned sound stimulus was given before, simultaneously with, or after the shock. Bernstein's results on backward conditioning are not in harmony with those of Switzer (**20**) and Wolfle (**23**; **24**). When the sound followed the shock in the C–R training, Bernstein states that "Only seven of the twenty subjects tested exhibited conditioned reactions, and the greatest amount of conditioning obtained in any subject was 8 percent" (**2**, 189). The sound was given .5 sec after the shock with 10 S's in the C–R training, and the percentage of conditioning was .8 using the Mean and 0 using the Median. The sound was given .9 sec after the shock with 10 other S's in the C–R training, and the percentage of conditioning was 1.6 using the Mean and 0 using the Median. It may be in order to raise the following questions: (1) Do these results show that backward conditioning occurred or that it did not occur? And (2), Is it not possible that the extremely small Mean percentages of eyelid reactions in the C–R trials, .8 and 1.6, were not due to a process of backward conditioning at all but to extraneous factors in the experimental situation which could not be readily controlled?

In view of the difficulty of controlling all conditions of an experiment and the probability that inner speech has some influence on almost all, if not all, types of motor learning, we

are inclined to favor the following three unproved but frequently and widely discussed propositions. (1) In training of the C–R type, if brain activity *A* begins and ceases, then there is a period of time, and finally brain activity *B* begins and ceases, there will be no tendency whatever for brain activity *A* to produce or lead to brain activity *B* in the future. (2) If brain activities *A* and *B* are present at the same time, that is, if the two organic processes overlap each other temporally, a connection may be formed from *A* to *B*, from *B* to *A*, or from *A* to *B* and from *B* to *A*. (3) In training of the C–R type, if brain activity *A* has ceased or has practically faded when brain activity *B* begins, there will be no tendency at all for brain activity *B* to produce or lead to brain activity *A* in the future. The experimental results on backward conditioning give some support to the last proposition.

(Manuscript received July 10, 1934)

References

1. Beritoff, J., Über die individuell-erworbene Tätigkeit des Zentralnervensystems bei Tauben, *Pflüger's Arch. f. d. ges. Physiol.*, 1926, **213**, 370–406.
2. Bernstein, A. L., Temporal factors in the formation of conditioned eyelid reactions in human subjects, *J. Gen. Psychol.*, 1934, **10**, 173–197.
3. Carr, H. A., & Freeman, A. S., Time relationships in the formation of associations, *Psychol. Rev.*, 1919, **26**, 465–473.
4. Cason, Hulsey, The conditioned pupillary reaction, J. Exper. Psychol., 1922, **5**, 108–146.
5. Cason, Hulsey, The conditioned eyelid reaction, J. Exper. Psychol., 1922, **5**, 153–196.
6. Cason, Hulsey, A note on the conditioned eyelid reaction, J. Exper. Psychol., 1923, **6**, 82–83.
7. Cason, Hulsey, The concept of backward association, *Amer. J. Psychol.*, 1924, **35**, 217–229.
8. Cason, Hulsey, General aspects of the conditioned response, *Psychol. Rev.*, 1925, **32**, 298–316.
9. Cason, Hulsey, Specific serial learning; a study of backward association, J. Exper. Psychol., 1926, **9**, 195–227.
10. Cason, Hulsey, Specific serial learning; a study of remote forward association, J. Exper. Psychol., 1926, **9**, 299–324.
11. Cason, Hulsey, The role of verbal activities in the conditioning of human subjects, *Psychol. Rev.*, 1934, **41**, 563–571.
12. Cohen, L. H., Periodicity of recovery during the refractory phase of the eyelid reflex, J. Exper. Psychol., 1932, **15**, 436–446.
13. Guthrie, E. R., Association as a function of time interval, *Psychol. Rev.*, 1933, **40**, 355–367.

14. HUMPHREY, GEORGE, *The Nature of Learning in its Relation to the Living System*, 1933.
15. KRESTOVNIKOV, A. N. (Experiments summarized in I. P. Pavlov's *Conditioned Reflexes*, tr, 1927, pp. 26–28.)
16. PAVLOV, I. P., *Conditioned Reflexes*, tr, 1927.
17. PAVLOV, I. P., *Lectures on Conditioned Reflexes*, tr, 1928.
18. PETERSON, JOSEPH, & ALLISON, L. W. Controls of the eye-wink mechanism, J. EXPER. PSYCHOL., 1931, **14**, 144–154.
19. PONDER, ERIC, & KENNEDY, W. P., On the act of blinking, *Quart. J. Exper. Physiol.*, 1927, **18**, 89–110.
20. SWITZER, ST. C. A., Backward conditioning of the lid reflex, J. EXPER. PSYCHOL., 1930, **13**, 76–97.
21. TELFORD, C. W., & THOMPSON, NORMAN, Some factors influencing voluntary and reflex eyelid responses, J. EXPER. PSYCHOL., 1933, **16**, 524–539.
22. WENRICK, J. E., The conundrum of the conditioned response, *Psychol. Rev.*, 1933, **40**, 549–559.
23. WOLFLE, H. M., Time factors in conditioning finger-withdrawal, *J. Gen. Psychol.*, 1930, **4**, 372–378.
24. WOLFLE, H. M., Conditioning as a function of the interval between the conditioned and the original stimulus, *J. Gen. Psychol.*, 1932, **7**, 80–103.

INFLUENCE OF SYLLABIC LENGTH AND RATE OF AUDITORY PRESENTATION ON ABILITY TO REPRODUCE DISCONNECTED WORDS LISTS[1]

BY STEPHEN WALLACE CALHOON

Ohio State University

The purpose of this study is to examine the effect of changes in the syllabic length of the words upon ability to reproduce words lists. Observations are also presented upon the effect of change in the rate of initial auditory presentation, when measured in terms of the number of words reproduced.

I. Method of Attack

Comparatively detailed report has already been made (**2**) of the experimental set-up by which the data of this study were secured. It will therefore be necessary only very briefly to review the method here.

By means of word lists of known difficulty, so selected as to be free from any apparent possibilities of associations, and pronounced by the experimenter in a manner carefully designed to eliminate variations in stimulational intensity, data were collected from 946 subjects, distributed quite evenly through the third to the twelfth grades inclusively, of seven different schools in six counties in central Ohio. Practice effects were carefully balanced and timing rigorously controlled and scrupulously checked. The experimental situation was consistently and gratifyingly free from extraneous environmental disturbances (**1**).

By this general method, data were collected for 4, 6, 8, 10, and (in Higher Grades) 12 one-syllable word lists; for 2, 4, 6, and 8 two-syllable word lists in Lower Grades; for 2, 4, 6, and 8 three-syllable word lists; and for 2, 4, and 6 four-syllable word lists in Higher Grades. Data were collected for situa-

[1] This research project was financed by the Payne Fund from money made available for the promotion of research on problems of radio education.

tions where words were pronounced at one-second, at two-second, and at three-second intervals. Standard times, experimentally determined, were allowed for the writing back of each list.

Results (1) show that the words used could be written correctly when memory is not a factor, by 56 percent of the third-grade pupils, while about 90 percent of the words were intelligibly written. Timing of the presentation of the lists was in error by an amount which varied within an Inter-Quartile Range of 5.55 percent, about a Mean of 3.00 percent, and the timing of writing periods was yet more accurate.

II. Treatment of Data and Results

The data thus collected were sorted, scored by standard scoring keys according to standard rules, assembled on general data assembly sheets, and retabulated into distributions both with regard to various types of lists presented at various rates, and also with reference to the Mean Grade-Ages of the subjects employed. From these numerous distributions, Arithmetical Means were extracted and tabulated. These Tables are presented complete at another place (1).

Evidence has already been reported (2) to show that the number of words reproduced is independent of excessive length of the lists pronounced. It has also been shown that two-word lists and four-word lists of all syllabic lengths at all rates, are too short to measure the extent of the word-span, a span which is manifestly present in every case. Similarly must we question, from a theoretical viewpoint, the sufficient length of six-word lists, though this is not practically demonstrated. But in the absence of more conclusive evidence regarding this list length, we have used it in the data of this report. However, two-word lists and four-word lists have been eliminated and general Means, without reference to list length, at each Mean Grade-Age level, have been computed from the specific Means of the various lists lengths at these various levels. These Means are reported in Tables I & I*a* of this paper. Table I*a* reports the average ages and the number of subjects from which these means were extracted.

Table II is a rearrangement of Table I in such a form that it becomes possible to observe the influence of syllabic length, when time is held constant.

III. Discussion and Conclusions

A. Reading of Tables

There are four sections to Tables I & I*a*, from left to right. The extreme left section presents results for one-syllable words. The next section presents results for two-syllable words, the next for three-, and the last for four-syllable words. In each section there are three columns of Means for words at one-second, two-second, and three-second intervals, respectively. Vertically the table represents variation with Mean Grade-Age of the subjects examined. Mean Grade-Age is tabulated above and Mean number of subjects tested below and in parentheses, in Table I*a*.

Therefore the quantity 1.7, after the quantity 8.9, in the three-second column of the first section of the table, signifies the fact that those in the third grade who wrote one-syllable words after the words were presented at the rate of one per three-seconds, were able, regardless of length of list presented,

TABLE I

Relation Between Syllabic Length, Rate of Presentation and Mean Number of Words Written

Grade	Syllabic Length											
	1			2			3			4		
	Rate of Presentation											
	1 Sec	2 Sec	3 Sec	1 Sec	2 Sec	3 Sec	1 Sec	2 Sec	3 Sec	1 Sec	2 Sec	3 Sec
3	2.1	2.6	1.7	1.8	2.0	1.5	1.7	2.1	1.4			
4	2.9	3.0	2.9	2.5	2.4	2.6	2.0	2.3	2.4			
5	3.1	3.1	3.3	2.9	3.1	2.8	2.7	2.9	2.6			
6	3.6	3.4	3.8	3.1	3.4	3.4	3.0	3.2	3.1			
7	3.5	3.7	4.1				3.0	3.1	3.6	2.3	2.3	2.8
8	3.8	4.0	4.4				3.2	3.4	3.8	2.6	2.5	3.1
9	3.9	4.2	4.0				3.5	3.4	3.2	2.8	3.7	2.1
10	4.0	4.2	4.6				3.8	3.6	4.0	2.9	3.5	3.3
11	5.0	5.0	4.4				4.2	4.2	4.2	3.6	3.9	3.6
12	4.4	4.7	4.9				4.2	3.9	4.4	3.5	3.7	3.8

TABLE I*a*

MEAN GRADE-AGE OF SUBJECTS AND NUMBER OF CASES FOR EACH MEAN OF TABLE I
(Ages in years; Number of Cases in Parentheses)

Grade	Syllabic Length											
	1			2			3			4		
	Rate of Presentation											
	1 Sec	2 Sec	3 Sec	1 Sec	2 Sec	3 Sec	1 Sec	2 Sec	3 Sec	1 Sec	2 Sec	3 Sec
3	9.1 (72)	9.0 (71)	8.9 (44)	9.1 (72)	9.0 (71)	8.9 (44)	9.1 (72)	9.0 (71)	8.9 (44)			
4	10.5 (88)	10.4 (90)	10.0 (59)	10.5 (88)	10.4 (90)	10.0 (59)	10.5 (88)	10.4 (90)	10.0 (59)			
5	11.1 (65)	10.8 (102)	10.8 (94)	11.1 (65)	10.8 (102)	10.8 (94)	11.1 (65)	10.8 (102)	10.8 (94)			
6	12.4 (52)	12.1 (90)	12.0 (93)	12.4 (52)	12.1 (90)	12.0 (93)	12.4 (52)	12.1 (90)	12.0 (93)			
7	13.4 (43)	13.5 (89)	13.5 (81)				13.4 (43)	13.5 (89)	13.5 (81)	13.4 (43)	13.5 (89)	13.5 (81)
8	14.2 (55)	14.1 (108)	14.0 (104)				14.2 (55)	14.1 (108)	14.0 (104)	14.2 (55)	14.1 (108)	14.0 (104)
9	15.3 (37)	15.1 (18)	15.5 (19)				15.3 (37)	15.1 (18)	15.5 (19)	15.3 (37)	15.1 (18)	15.5 (19)
10	16.3 (81)	16.2 (79)	15.8 (45)				16.3 (81)	16.2 (79)	15.8 (45)	16.3 (81)	16.2 (79)	15.8 (45)
11	16.9 (81)	16.9 (88)	16.9 (26)				16.9 (81)	16.9 (88)	16.9 (26)	16.9 (81)	16.9 (88)	16.9 (26)
12	18.1 (33)	18.0 (42)	18.1 (13)				18.1 (33)	18.0 (42)	18.1 (13)	18.1 (33)	18.0 (42)	18.1 (13)

to reproduce a Mean of 1.7 words. Those in the fourth grade, who reproduced one-syllable words after they had been pronounced at the rate of one per three seconds, had a Mean Grade-Age of 10.0 years and were able to reproduce a Mean of 2.9 words. The number of individual cases from which these Means are derived are represented in the upper portion of each square in the Mean-Grade-Age column. Each Mean is the average of from one to four list lengths.

Table II is divided into three portions. The first presents all data for various syllabic-lengths of words presented at the

TABLE II

RELATION BETWEEN INTERVAL, SYLLABIC LENGTH AND MEAN NUMBER OF WORDS WRITTEN

Words at															
1 Second						2 Seconds						3 Seconds					
Mean Gr–Age	No. of Cases	Syllabic Length				Mean Gr–Age	No. of Cases	Syllabic Length				Mean Gr–Age	No. of Cases	Syllabic Length			
		1	2	3	4			1	2	3	4			1	2	3	4
9.1	72	2.1	1.9	1.7		9.0	71	2.6	2.1	2.1		8.9	44	1.7	1.5	1.4	
10.5	88	2.9	2.5	2.0		10.4	90	3.0	2.5	2.3		10.0	59	2.9	2.6	2.4	
11.1	65	3.1	2.9	2.7		10.8	102	3.1	3.1	2.9		10.8	94	3.3	2.8	2.6	
12.4	52	3.6	3.1	3.0		12.1	90	3.4	3.4	3.2		12.0	93	3.8	3.4	3.1	
13.4	43	3.5		3.0	2.3	13.5	89	3.7		3.1	2.4	13.5	81	4.1		3.6	2.8
14.2	55	3.8		3.2	2.6	14.1	108	4.0		3.4	2.5	14.1	104	4.4		3.8	3.1
15.3	37	3.9		3.5	2.8	15.1	18	4.2		3.4	3.7	15.5	19	4.0		3.2	2.1
16.3	81	4.0		3.8	2.9	16.2	79	4.2		3.6	3.5	15.8	45	4.6		4.0	3.3
16.9	81	5.0		4.2	3.6	16.9	88	5.0		4.2	4.0	16.9	26	4.4		4.2	3.7
18.1	33	4.4		4.2	3.5	18.1	42	4.7		3.9	3.7	18.1	13	4.9		4.4	3.8

rate of one word per second. The second portion presents data for all words at two-second intervals and the third for all words at three seconds. Each section has four columns for one-syllable, two-syllable, three-syllable, and four syllable words. Each Mean in the individual rows of this table is for the same Mean Grade-Age. From one to four list averages are included in each Mean. This Table is read much like Table I.

Now let us turn to an examination of these tables for tendencies which they present.

B. Irregularities

1. Irregularities in the data at the ninth-grade level, for various intervals and for Mean Grade-Ages of 15.3, 15.1, and 15.1, are manifestly due to sampling errors resulting from insufficient population, since there were only 37, 18, and 19 cases in each instance, respectively, as shown by Table I*a*, and an examination of the records shows that these cases were all taken from a school in a backward community.

2. Irregularities at the Eleventh-Grade level are also due to a biased population resultant from the testing of two classes of French students in a larger school system where French was an elective course; 53 of the 95 Juniors tested were French pupils in these two classes. And we know that the brighter pupils elect French. See Terman's (**3**) discussion on this point.

3. When correction is allowed for the irregularities thus introduced, certain relationships very clearly appear, especially if the data of Tables I and II be plotted (**2**). But an examination of the Tables themselves reveal the same tendencies, of course.

C. Rate of Presentation and Ability to Reproduce

When the nature of lists is held constant and rate of presentation is allowed to vary (Table I), a group of rather heterogeneous tendencies are observed which seem to be best described by the rule: *When interval between words increases, the slope of the curves, in the lower age levels, becomes more acute*

with a more and more prominent negative acceleration at the higher ages.

To prove this, let us look at Table I. We see that, for one-syllable words at one-, two-, and three-second intervals, the three-second interval yields the lowest initial Mean but quickly climbs, so that, by the fifth-grade Mean-Age, it has become superior to both the one-second and the two-second intervals. It unquestionably maintains this superiority until after the tenth-grade, when the relationship, described by the results of this experiment, becomes less clear. But it is probable that the same relationship, only to a less degree, perhaps, continues. Also, we can see that the two-second interval yields Means consistently greater than the one-second interval but also consistently less than the three-second interval, at least throughout the age range from 10 to 16.5 years. And similarly this relationship may continue throughout the Higher-Grade levels to a lesser degree. The irregularity of the sixth grade Mean Grade-Age point, for words at one second, probably finds its origin in the variable associability of word lists pointed out in a previous paper (2) and is therefore not due to any fundamental age function. Thus we see that the data of Table I, for one-syllable words, clearly substantiate the rule given above.

Similarly do the data for three-syllable words show the same general relationship throughout, if we make proper allowance for the influence of variable associability between words at one- and at two-second intervals. And the data for two-syllable words in the Lower Grades show the same relationship. The relationship in four-syllable presentations, for the Higher Grades, is not as clear, but, at least, no contradictory tendency is very prominently present.

It has been suggested that the difference in results between these rates of presentation are not large enough to be of very much significance, but let us point out the fact that the data of Table I, for one-syllable words, are gathered from four samplings of reproductions from each individual in a series of populations varying in size from 13 to 114, with a Mean of 64.7 cases. Therefore we must observe that, though the

effects of biased sampling are continued, individual variations, and, for the most part, variations in associability of lists presented, should be more completely stabilized. Consequently a general relationship seems to be here pointed out, and either substantiated or else not discounted by all the data of the experiment. It is the belief of the present investigator that further yet more carefully controlled experimentation should even more clearly demonstrate this relationship.

D. Syllabic Length and Ability to Reproduce

Table II presents the results in such a manner that rate of presentation is held constant while syllabic length is allowed to vary.

Here, any heterogeneous tendencies that might before (Table I) have been present, are entirely removed and the organization of relationships which presents itself immediately becomes very clear.

An examination of the Mean number of words reproduced for various syllabic lengths reveals an orderly, descending magnitude with ascending syllabic length at every Mean Grade-Age point examined, for words at one second 2.1, 1.9, 1.7, 2.9, 2.5, 2.0, 3.1, 2.9, and 2.7, for one-syllable, two-syllable, and three-syllable words for Mean Grade-Ages 9.1, 10.5, and 11.1, respectively, are the orders in which these Means run. And the same relationship continues with two slight exceptions throughout the Age range for words at two-second intervals and also, but without exception, for words presented at intervals of three seconds.

Consequently we may quickly and unquestioningly conclude that, *as time remains constant, increased syllabic length of words auditorily presented, decreases the number of words written in an inverse proportion.* The exact mathematical relationships we do not venture to predict, because of the varying associability of the lists employed (**2**).

III. Summary

From data reported in this study, it quite conclusively appears that

1. When the interval between words increases, syllabic length remaining constant, the slope of the curves of ability to reproduce words auditorially presented, becomes more acute at ascending age levels, with a more and more prominent negative acceleration as the eighteenth year is approached. This negative acceleration first appears about the thirteenth year.

2. Ability to reproduce words auditorially presented is in an inverse relation to their syllabic length when rate of presentation remains constant. The mathematical ratios are not yet known.

(Manuscript received June 11, 1934)

References

1. Calhoon, S. W., *Studies In Auditory Impressionability*, M.S. Bound and on file with The Psychological Exchange Library, New York City, and also with the Bureau of Educational Research Library, College of Education, Ohio State University, Columbus.

2. Calhoon, S. W., Influence of Length of Lists Upon Ability Immediately to Reproduce Disconnected Words Series Auditorily Presented. *Jour. of Exp. Psychol.* 1934, **17**, 723-738.

3. Terman, L. M., *Mental and Physical Traits of 1000 Gifted Children.* (Vol. I. *Genetic Study of Genius*—Terman), Stanford University: Stanford University Press, 1925.

A COMPARISON OF ABILITY TO REPRODUCE ONE-SYLLABLE WORDS AND DIGITS AUDITORIALLY PRESENTED [1]

BY STEPHEN WALLACE CALHOON

Ohio State University

It appears that a careful comparison of the ability to reproduce various types of disconnected items auditorially presented, through a large age-range, has never been made. It is the purpose of this article to present such a comparison. It is a report of the comparison of developing ability to reproduce one-syllable words and of ability to reproduce disconnected numbers or digits, auditorially and comparably presented.

The Method

This study was conducted as a unit in a larger study upon the problem of Auditory Impressionability. The larger study (3) was made from data collected in five experimental sessions with each group of subjects employed, each session lasting fifteen minutes, after a rest period of at least thirty-minutes had elapsed between. The data reported here were always collected at the fifth session. Three of the four other sessions were employed in the reproducing of word lists of varying lengths and syllabic constitution, and the fourth was employed in the writing of words pronounced after the style of an ordinary written spelling lesson. Consequently, since each memory test had either eleven or twelve lists, practice in reproducing lists had been experienced by everyone who took this fifth test to the extent of either 33 or 36 lists.

The 50 words pronounced in the spelling test given first, were selected by random sampling methods from Tidyman's (8) third-grade lists, the same list from which the one-syllable words used in this study were selected. And a subsequent examination of the ability of the subjects to spell these words (3) reveals the fact that in 90 percent of the cases third graders could write the words intelligibly, though the laboriousness of writing and the grotesqueness of spelling, in some cases, leads us to suppose that the true curves would be, absolutely, a little higher, for third-graders—and, perhaps, for fourth-graders. Others had no difficulty in writing the words pronounced.

From the words of Tidyman's third-grade list, one list each of 4, 6, 8, and 12, one-syllable words was selected at random, care being exercised to see that no alphabetical sequence occurred and that no obvious possibility for association between any two successive words was present. Lists of 3–10 digits, the lists employed by Humpstone (7), and so selected as to have no consecutive order, no repetitions, except in the ten-digit list, no zeros and no familiar or significant combinations such as '365' or '144'

[1] This research project was financed by the Payne Fund from money made available for the promotion of research on problems of radio education.

were selected and the lists of digits and words intermixed by chance. These thirteen lists were then incorporated into the last test of the series of the major experiment.

The test was introduced with the following self-explanatory directions:

> "This test contains some lists of words and some lists of numbers. I shall tell you which each list will be, words or numbers. You are to write these lists in the squares as you have done before—words and numbers alike—*except*, that you are to write these lists, as nearly in the order that I give them as possible. That is, if I say '1–2–3–4—now,' you are to write the numbers as nearly in the order 1–2–3–4 as you can. If you cannot remember the order in which the words or numbers were given, but do remember the words or numbers themselves, write them down anyway—but give them as nearly in the right order as possible. Pencils up! Mouths closed! This list will be numbers! Ready":

The lists were then given in the following order:

(1) 3 digits: 4–9–7.
(2) 12 words: hand, day, card, did, bone, cross, hat, fold, and, cry, dive, feed.
(3) 4 digits: 8–2–6–4.
(4) 4 words: laugh, row, tuck, church.
(5) 8 digits: 1–4–7–2–9–6–3–8.
(6) 6 words: rice, tongue, nest, dog, knee, dust.
(7) 9 digits: 3–7–5–2–8–6–4–9–1.
(8) 5 digits: 4–1–8–6–3.
(9) 7 digits: 6–9–3–7–1–5–8.
(10) 8 words: out, string, hear, space, sent, build, grade, spin.
(11) 6 digits: 7–2–4–9–5–3.
(12) 10 words: walk, say, eight, felt, long, sent, weave, brown, gone, neat.
(13) 10 digits: 9–2–7–4–1–6–9–3–5–8.

The lists were given in the same order every time, under the assumption that the

TABLE I

RELATION BETWEEN MEAN GRADE-AGES AND NUMBER OF WORDS AND DIGITS WRITTEN
(Words at One Second)

Nature of Lists	Number of Words and Digits Written									
3 Digits	2.7	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
4 Digits	3.4	4.0	3.9	4.0	4.0	4.0	4.0	4.0	3.9	4.0
5 Digits	3.9	4.4	4.6	4.7	4.8	4.9	5.0	4.9	4.9	4.9
6 Digits	4.0	4.8	4.8	5.0	5.4	5.5	5.5	5.4	5.4	5.4
7 Digits	4.1	5.0	5.4	5.6	5.8	6.0	6.0	5.8	6.5	6.0
8 Digits	4.3	5.3	6.0	6.1	6.4	6.7	6.3	6.6	6.9	6.8
9 Digits	4.5	5.5	5.9	3.3	6.7	7.0	6.8	6.5	7.5	6.9
10 Digits	4.7	5.0	6.6	6.7	6.8	7.4	7.1	6.9	7.3	6.4
4 Words	1.8	3.1	3.2	3.5	3.7	3.8	3.9	3.9	4.0	3.8
6 Words	2.3	3.1	3.3	3.9	3.9	4.2	4.6	4.1	5.0	4.4
8 Words	2.1	3.1	3.1	3.3	3.8	4.2	4.7	3.9	4.8	4.6
10 Words	2.3	3.9	3.0	3.5	4.0	4.2	4.8	4.6	5.8	4.5
12 Words	2.1	3.0	3.0	4.1	4.4	5.0	5.4	4.4	5.1	4.5
Mean Grade-Age	8.1	9.7	10.5	12.0	13.5	13.8	14.9	16.6	17.0	18.0
Number of Subjects	14	21	59	55	56	72	16	14	16	8

TABLE II

RELATION BETWEEN MEAN GRADE-AGES AND NUMBER OF WORDS AND DIGITS WRITTEN
(Words at Two Seconds)

Nature of Lists	Number of Words and Digits Written								
3 Digits	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
4 Digits	3.7	4.0	4.0	3.9	4.0	4.0	3.9	4.0	4.0
5 Digits	4.1	4.5	4.8	4.8	4.0	4.8	5.0	4.9	4.9
6 Digits	4.3	4.7	5.0	5.7	5.4	5.5	5.2	5.5	5.3
7 Digits	4.4	4.9	5.9	5.6	5.9	6.1	5.1	6.2	5.7
8 Digits	5.3	5.7	5.9	6.3	6.8	6.6	6.8	7.2	6.7
9 Digits	5.2	5.7	6.0	6.8	6.6	7.0	7.3	6.9	6.6
10 Digits	5.8	5.6	6.4	7.1	6.5	7.1	7.0	6.6	5.9
4 Words	1.7	2.7	3.0	3.8	3.9	3.7	3.9	3.8	3.9
6 Words	2.3	2.7	3.2	4.0	4.2	3.9	4.3	4.6	3.9
8 Words	1.6	2.6	3.1	3.8	4.0	3.9	4.0	4.0	4.1
10 Words	2.3	2.9	3.1	4.0	3.9	4.1	4.5	4.1	4.4
12 Words	2.6	3.3	4.1	4.6	5.4	5.1	4.7	5.1	4.8
Mean Grade-Age	9.4	9.9	11.4	12.2	13.4	14.2	15.5	15.9	16.8
Number of Subjects	15	21	18	17	10	19	19	13	10

TABLE III

RELATION BETWEEN MEAN GRADE-AGES AND NUMBER OF WORDS AND DIGITS WRITTEN
(Words at Three Seconds)

Nature of Lists	Number of Words and Digits Written									
3 Digits	3.0	3.0	3.0	2.9	3.0	3.0	3.0	3.0	3.0	3.0
4 Digits	3.9	3.9	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
5 Digits	4.7	4.6	4.6	4.6	4.8	4.9	5.0	4.9	5.0	5.0
6 Digits	4.8	5.0	4.6	4.9	5.1	5.6	5.5	5.7	5.8	5.8
7 Digits	4.8	4.8	5.1	5.1	6.3	6.2	6.0	6.1	6.5	6.5
8 Digits	5.7	6.0	6.2	5.7	6.5	6.6	6.3	7.0	7.2	7.4
9 Digits	5.8	5.7	6.0	6.1	7.2	6.7	6.8	7.2	7.6	8.0
10 Digits	6.2	6.0	6.0	6.1	7.6	7.5	7.1	7.9	7.3	8.1
4 Words	2.7	2.9	3.1	3.1	3.8	4.0	3.9	3.9	4.0	4.0
6 Words	3.0	3.1	3.5	3.9	3.9	4.3	4.6	4.8	5.1	4.9
8 Words	2.8	2.8	3.1	3.1	3.4	4.1	5.0	4.8	4.8	5.6
10 Words	2.7	2.8	3.0	3.0	3.9	4.6	4.8	4.9	5.3	5.5
12 Words	3.2	3.8	4.1	3.4	4.6	5.0	5.4	6.0	5.4	6.5
Mean Grade-Age	9.2	10.9	11.1	12.9	13.4	14.3	14.9	16.5	16.9	17.5
Number of Subjects	41	52	31	14	18	23	16	45	71	30

33–36 lists of practice which all pupils had experienced, would have accomplished the improvement in ability to reproduce for the day. Constant times were allowed for the writing of the lists. The lists were pronounced at one-second, two-second and three-second intervals, in various classes, so that the numbers of pupils reported in Tables I–III, for various Mean-Grade-Ages, were, at varying rates of presentation, tested.

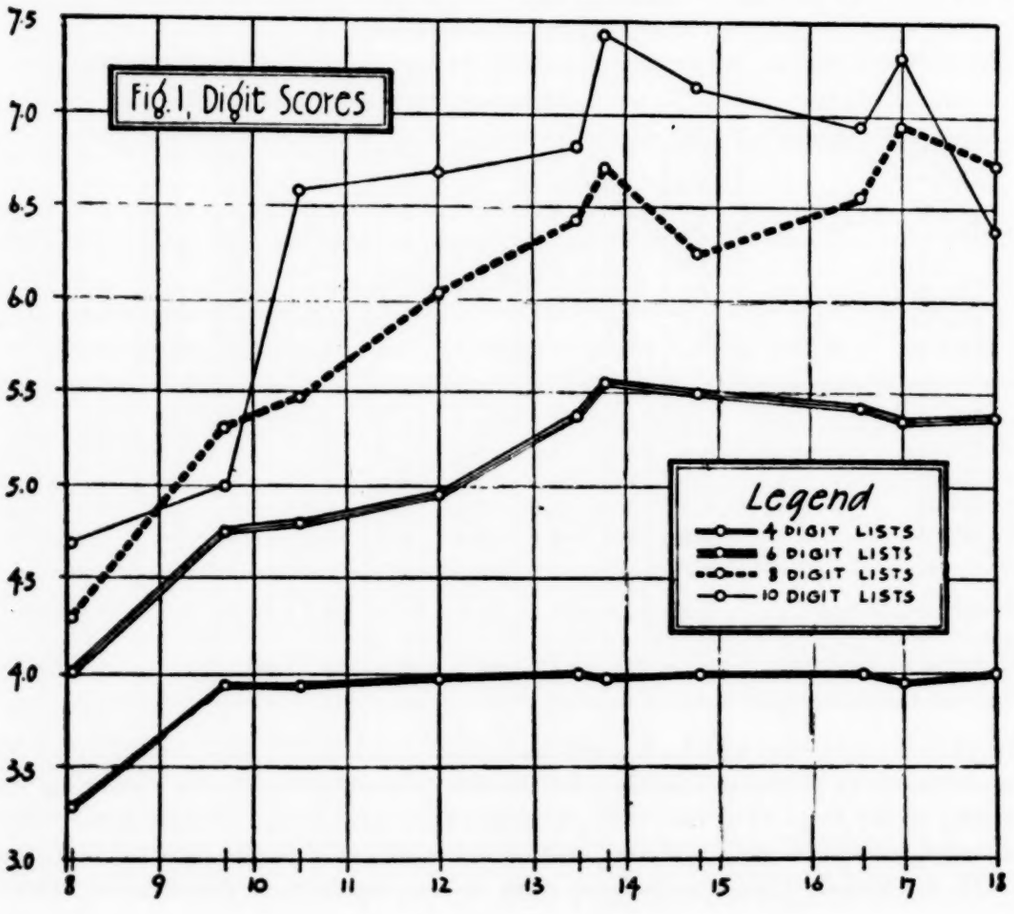
Words and digits were pronounced without accent, in a quiet but firm, monotonous voice. The timing of the pronunciation of lists was done by rhythms secured from the second hand of the experimenter's watch and practiced throughout the writing periods for the various lists, with eye upon the watch. The pronunciations were recorded on a dictaphone record and all errors to the half-second for entire lists for both pronouncing and writing were later tabulated and studied (3). This study of timing errors reveals a Mean error of 3 percent with an interquartile range of 5.55 percent for the timing of pronunciations and less error for timing the writing of the lists. All children were generally through writing before close of the writing period. Environmental disturbances were satisfactorily absent (3).

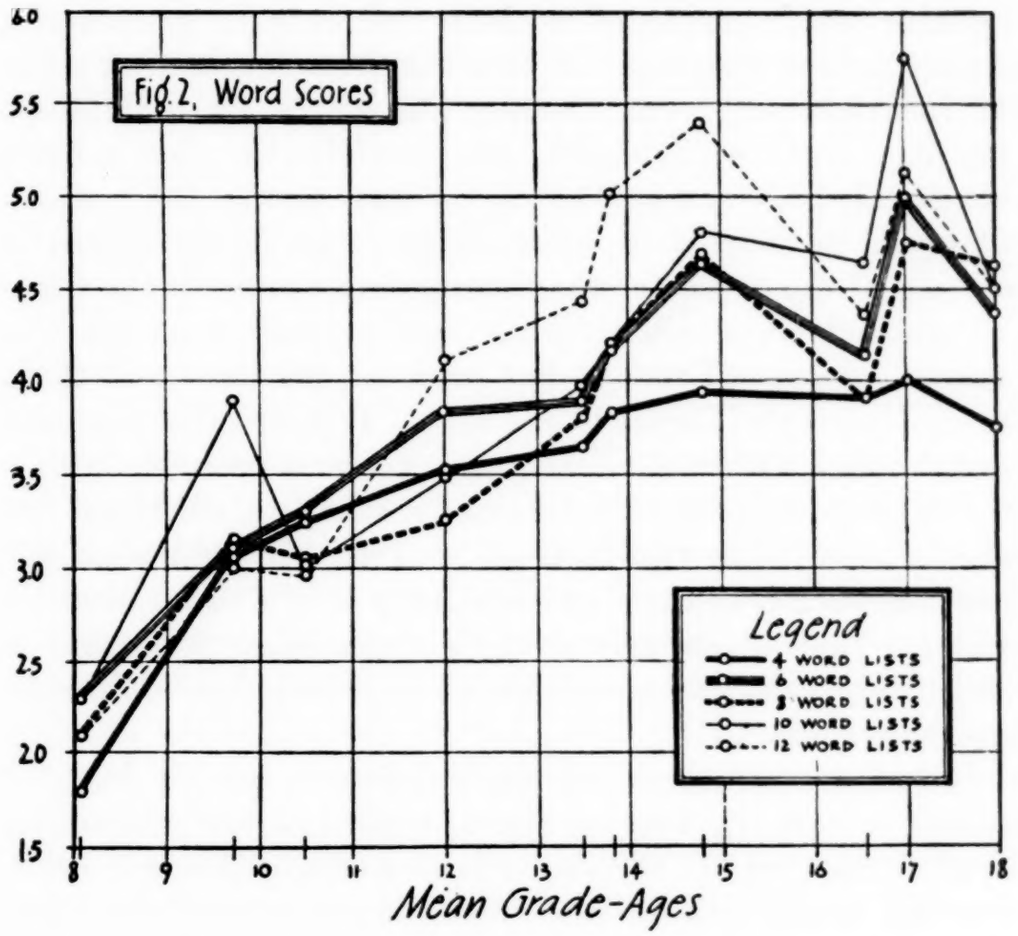
Consequently the data of this study seems to have been collected from a situation well-controlled environmentally, by a measuring instrument almost entirely adapted to the mechanical requirements of the problem and under a timing scheme so directed that a mean error of only 3 percent appeared. Practice effects were also controlled and the attitude of the pupils, one of almost universal competition, was built up through four preceding fifteen-minute tests.

The Results

By means of this set-up, the data of Tables I, II, and III, samples of which are graphically represented in Figs. 1 and 2, were gathered.

Tables I, II, and III are simple and seem quite self-





explanatory. They are all read in the same manner. For example, the number 2.7 in Table I, above 'Mean Grade-Age 8.1' and after '3-digits,' signifies the fact that 14 third-grade pupils with a Mean Age of 8.1 years, at the nearest birthday, reproduced a Mean of 2.7 digits of a three-digit list pronounced. In a like manner, the quantity 1.4, after the notation, '4-words,' in the same column, means that 14 third-graders with a Mean age of 8.1 years, reproduced a Mean of 1.8 words from a four-word list. And thus are all the quantities interpreted. Figs. 1 and 2 are plottings of the data of Table I, with Mean numbers of units reproduced for abscissae and Mean Grade-Ages for ordinates.

Discussion

1. *For Words*

Evidence has already been brought forth (4), definitely to show that the simple facts of random sampling of the one-

syllable words employed without reference to grammatical nature, and arrangement in lists free from mechanical marks of identification, even though the words be individually familiar and mechanically reproducible by the subjects examined, does not guarantee, by any means, the presence of equal coefficients of associability. Hence we cannot be certain of the cause for the outstanding nature of the list of 12 one-syllable words showing itself somewhat at the one-second interval (Table I) but more at the two- and three-second intervals (Tables II and III). That it is due to greater associability within the list, however, seems probable both in the light of the work of the major experimental disclosure and also in view of the two facts (1) that neither eight-word lists nor ten-word lists show any tendency in the same direction and (2) that the phenomenon is much more prominent at longer intervals between words where association could more easily be made.

If we are to accept this explanation, it quickly becomes apparent from a glance at Fig. 2, typical of the relationship at all times, that, for the one-syllable words of this test, excessive list length has no consistent influence upon the Mean number of words reproduced, for we see that the curves of all the list lengths fall very closely together, crossing and recrossing each other in such a manner as to denote no consistent difference between them. This confirms the conclusion previously brought forth (4). And this same relationship is confirmed also by the data of Tables II and III, the results when words are presented at two-second and at three-second intervals, respectively, for we quickly see, from an examination of Tables I, II, and III, that there is no consistent difference between list lengths as indicated by the Mean number of words reproduced, at any interval. In Table IV, we find the summary of these relationships.

The number 3.4, in the column marked '1-Second' and the row marked '4 Words' means that, for four-word lists, at one-second intervals, a mean of 3.4 words was reproduced by all the subjects examined. Similarly, for six-word lists at one second, a mean of 3.8 words was reproduced by all subjects examined, etc., throughout the table.

TABLE IV

SUMMARY OF INFLUENCE OF LIST LENGTH ON MEAN NUMBER OF WORDS WRITTEN

No. Pronounced	Interval		
	1″	2″	3″
4 Words	3·4	3·4	3·5
6 Words	3·8	3·7	4·1
8 Words	3·7	3·5	3·9
10 Words	4·0	3·7	4·1
12 Words	4·0	4·4	4·7

We see, from this Table that the four-word list gives a lower Mean for every rate. This is, no doubt, explained by the fact that the Mean is approaching 4.0 as a limit and is therefore artificially suppressed by the influence of this limit. In other words, the four-word list is too short to measure the span accurately. This is borne out, graphically, in both the Figures presented and would be as strikingly manifest if other data were plotted. As for the remaining Means of Table IV, for six-, eight-, ten-, and twelve-word lists, fluctuations in value are comparatively slight, are for the most part inconsistent, and are explainable either in terms of variation of the coefficients of associability of the individual lists or by considering the errors of sampling in population studied.

2. *For Digits*

An examination of the data for digits under entirely comparable conditions, however, shows a striking difference. In Fig. 2 we have a graphical representation of the Mean number of digits reproduced from presentations at one-second intervals. This figure is comparable, therefore to Fig. 1, and it is typical, too, of the arrangement of the data for digits at two-second and at three-second intervals (Tables II and III). This is demonstrated by the summary, of Table V, read in the same manner as Table IV.

Here we clearly see the increase from list length to list length for digits at all rates. But further examination of the data reveals the fact that there is no doubt but that the two-, four-, five-, and even the six-digit curves, for ages above 14 years (Fig. 2), at least, are artificially suppressed by the limit-

TABLE V

SUMMARY OF INFLUENCE OF LIST LENGTH ON MEAN NUMBER OF DIGITS WRITTEN

No. Pronounced	Intervals		
	1″	2″	3″
3 Digits	3.0	3.0	3.0
4 Digits	3.9	3.6	4.0
5 Digits	4.7	4.0	4.8
6 Digits	5.1	4.6	5.3
7 Digits	5.6	5.5	5.7
8 Digits	6.1	5.7	6.5
9 Digits	6.0	6.4	6.7
10 Digits	6.5	5.8	7.0

ing length of the lists. This is obvious in all rates of presentation (Tables I–III) and appears at an even earlier age in the two-second rate (Table II). But there is, beside this, a negative acceleration in the increase of the Mean from list-length to list-length, at all intervals, which may indicate the presence of a limiting digit-span, measurable impartially throughout the age range studied, only by lists of digits yet longer than ten. In other words, as we read down the columns of values for the one-second, two-second, and three-second intervals, we find that the excess of each succeeding Mean over the preceding one becomes less. Hence it may be that even longer lists than ten digits are necessary to measure impartially the digit span, through all the age range studied.

That this increase from list-length to list-length is not due to more practice at succeeding lengths is clearly demonstrated by an examination of the order in which the various length lists were given (three-digits, four-digits, eight-digits, nine-digits, five-digits, seven-digits, and ten-digits). And it is also shown by the fact that either 33 or 36 lists for practice always preceded the test, while Humpstone (**4**), and Young (**6**), have demonstrated the fact that practice can only improve by about .6 of a digit and does not improve scores after eight practice lists have been written.

That this arrangement is further, not a chance one is demonstrated by the fact that it is confirmed by all the data collected from more than 800 subjects from seven different schools in six different counties in Central Ohio.

And finally, the fact that a list length yet longer than ten digits is necessary to measure impartially the digit span is shown by an examination of Fig. 2, where we find the ten-digit curve higher than the eight-, by a very considerable amount, at almost every point and at every interval to an important degree.

Consequently, it becomes apparent that work like that of Bolton (**2**) of Woolley and Fisher (**9**), of Gates (**5**), and of Guilford and Dallenback (**6**), is not free from the influence of limiting list length and must, as a result, yield curves, suppressed, at least in some portions.

3. *Words and Digits*

We observe, further, that there is a general tendency, at all ages, and at all rates, to reproduce a larger number of digits than of words by an amount which exceeds 2.7 units. To demonstrate this, we must point out, first, one or two other considerations.

An examination of Tables IV and V shows that, for both words and digits, lists presented at the two-second interval are, apparently reproduced with greater difficulty than either at one-second or at three-second intervals. But this is not a justifiable conclusion in light of the fact that the data for lists at two-seconds was all collected from one school, located in a comparatively poor and backward community, from a Mean number of 15.7 pupils per grade while the data for one-second, and for three-second intervals was gathered from three schools each, in better communities, with an average of 35 pupils per grade, for lists at one-second and of 34.1 pupils per grade for lists at three-second intervals. When we make due allowance for this sampling error we see why Means for two-second intervals should be smaller than for either of the other rates.

Now, if we take the Means for all reproductions of words, eliminating the four-word lists because they are too short and the twelve-word lists because they seem too easy, and compare the Means with the Mean number of digits reproduced from ten-digit lists previously demonstrated to give the nearest to the true Mean, we produce the results represented in Table VI:

TABLE VI

SUMMARY OF THE DIFFERENCES BETWEEN MEAN NUMBER OF WORDS AND OF DIGITS REPRODUCED FOR VARIOUS SCHOOL GRADES AND RATES OF PRESENTATION

School Grade	Difference Between Digits and Words At:			Average Difference
	1″	2″	3″	
3	1.5	2.1	3.4	2.3
4	1.7	2.9	3.1	2.6
5	3.5	3.3	2.8	3.2
6	2.9	3.1	2.7	2.9
7	2.9	2.5	3.9	3.1
8	3.2	3.1	3.2	3.2
9		2.7	2.4	2.6
10	2.7	2.4	3.0	2.7
11	3.1	1.8	2.3	2.4
12	1.9		2.8	2.3
Av. Dif.	2.6	2.7	3.0	2.7

In the second column of this table, under the caption 'One Second,' we have arranged, in order of school grade examined, the difference in units reproduced between the Means for six-, eight-, and ten-word lists and the Mean for the ten-digit list, of words and digits pronounced at one-second intervals. In every case the digit Mean is the larger. Similarly does column three express the difference for words and digits at two-seconds. Column five is the average of Columns 1, 2, and 3, and gives the Mean difference between number of words and digits reproduced.

We find, from this table that the Mean difference between words and digits is 2.7. But we have already shown that the true Mean for number of digits reproduced must be consistently higher than the Mean recorded for ten-digit lists, because of the limiting nature of the length of the ten-digit list. Consequently the difference between words and digits reproduced must be greater yet than 2.7.

But there yet is another relationship to be pointed out. As the interval between words and digits increases, there is more of an acceleration in the Mean number of digits reproduced than in the Mean number of words. This can be verified by a comparative study of Tables I, II, and III and it explains the increasing difference between the two Means at the lower grade levels, with reference to increasing interval

between the items of the lists (Table VI). On the other hand, negative acceleration also appears more quickly in the digit curves with reference to age (Tables I, II, and III). This explains the fact of the largest difference between the two spans from about the fifth grade through to the eighth (Table VI) and explains also the diminishing difference between the curves at higher grade levels (Table VI).

In other words, it appears that there is an advantageous influence of increased interval upon the number of digits which the younger pupils can reproduce which quickly reaches its greatest rate of development and then falls away to the horizontal plane more quickly while the reproduction of words is less influenced by increased interval and consequently the curves are more gradually and evenly pitched.

This is not at disagreement with the work of Bergström (1), who found digits to be less influenced by increased time than meaningful materials, for his work was conducted at the college level where more or less constant levels of ability with reference to age, for reproducing digits, at least, had been obtained. In fact, the contrasting degrees of negative acceleration between words and digits would tend rather, to substantiate his work; for the increase of negative acceleration argues a stabilization and maturity of ability which could not as easily be influenced by changes of rate at the college level. On the other hand, word spans are yet sensitive to alterations, for negative acceleration has not yet nearly as prominently appeared.

Conclusions

Therefore we may quite definitely conclude from the evidence of this study:

1. That in confirmation of work previously reported, the number of one-syllable words reproduced after a short interval of time is largely, if not altogether, uninfluenced by excessive length of the lists auditorily presented.

2. That this independence is consistently present throughout the age range studied.

3. That the digit span is longer consistently throughout the age range studied by more than 2.73 units.

4. That there is a differential relationship between ability to reproduce one-syllable words and ability to reproduce digits, which manifests itself in earlier acceleration in developing ability for digits, with a resultant earlier maturity of span.

5. That ten-digit lists are too short impartially to measure, statistically, the digit-span throughout the age range of this experiment.

6. That data reported by other investigators, which has not taken the necessity of this longer list length into account, cannot be expected to have described the true extent of the digit span.

(Manuscript received June 11, 1934)

References

1. Bergström, J. A., Effects of changes in the time variables in memorizing, together with some discussion of the techniques of memory experimentation, *Am. J. Psychol.*, 1907, **18**, 206–38.
2. Bolton, T. L., The growth of memory in school children, *Am. J. Psychol.*, 1892, **4**, 362–80.
3. Calhoon, S. W., *Studies in Auditory Impressionability*, M.S. Bound and on file with The Psychological Exchange Library, New York City, and also with the Bureau of Educational Research Library, College of Education, Ohio State University, Columbus, Pp. 75.
4. Calhoon, S. W., Influence of length of lists upon ability immediately to reproduce disconnected word series auditorially presented, *Journ. of Exp. Psychol.*, 1934, **17**, 723–738.
5. Gates, A. I., The mnemonic span for visual and auditory digits, J. Exper. Psychol., 1916, **1**, 392–403.
6. Guilford and Dallenbach, The determination of memory span by the method of constant stimuli., *Am. J. Psychol.*, 1925, **36**, 621–28.
7. Humpstone, H. J., *Some Aspects of the Memory Span—A Study in Association*, 1917: 10.
8. Tidyman, W. F., *Survey of the Writing Vocabularies of Public School Children in Connecticut.*, Teachers' Leaflet No. 15, U. S. Dept. of Interior, Office of Education, Washington, D. C.
9. Wooley, H. T., and Fisher, C. R., Mental and physical measurements of working children, *Psychol. Monog.*, 1914, **18**, 124.
10. Young, M. H., A comparative study of audito-vocal digit spans, *Psychol. Clinic.*, 1928, **17**, 170–83.

AN EXPERIMENT ON ORIENTATION IN STYLUS MAZE LEARNING

BY HENRY N. PETERS AND LOUISE McLEAN

University of Missouri

Dashiell's study of directional orientation in rats with an open-alley maze, which offered many equidistant pathways to the goal, indicates conclusively that (1) orientation toward the goal and its general direction is dominant over (2) orientation to local cues encountered in the pathway.[1] The striking result of this experiment was the finding that a rat tends to follow different, correct pathways on successive trials. This result is in harmony with an increasing number of experimental observations reported by Lashley[2] and others, indicating the non-serial habit nature of rodent maze learning. The maze habit with rats is definitely not a chaining together, or associative connecting, of fixed part-movements.

A large amount of work has been done with human stylus maze learning in experimental investigations of many phases of learning. There is no experimental evidence, however, upon which to base a conclusion concerning the relative importance of the two types of orientation distinguishable in maze learning with animals. An answer to this problem would have important theoretical implications for (1) human habits in general, (2) interpretation of the wealth of data in the literature on stylus maze learning, and (3) comparison of motor learning in human beings and in animals.

The immediate problem of the present experiment was: In stylus maze learning, if the human subject is presented with many equidistant pathways to the goal, will he fixate one of them or will he vary his choice of paths as the rat does?

The maze was made of wood and modeled after Dashiell's[3] in which two sets of criss-cross, parallel pathways offered twenty different equidistant (shortest) routes to the goal

and a very large number of longer ones. The pathways were ½ inch wide and ¼ inch deep. Parallel pathways were 2½ inches apart. Two culs-de-sac, the purpose of which was to eliminate the possibility of running the stylus along the outer wall of the outside pathways, were placed on each side, 1½ inches from a corner, making a total of eight culs-de-sac. Twenty-two college students, fourteen women and eight men, were used as subjects, fifteen of whom were inexperienced at maze learning.

The conventional stylus maze procedure was followed. The subject sat at the side of a table with the maze directly in front of him, but shielded from his view by a black cloth screen. One set of the maze paths was parallel with the side of the table. The starting point was the near left-hand corner; the goal, which was an exit from the maze, was at the far right-hand corner. The customary maze instructions were read to the subject. He was told that the criterion of learning was ten successive perfect trials and that the experimenter would indicate when he had reached it. Twenty-four hours after completion of learning, twenty-one of the subjects relearned the maze.

The experimenter recorded the exact pathway followed on each trial by means of numbers representing the segments of the open-alleys between two intersections. Any false turn, one leading to a round-about pathway, or any entrance into one of the eight culs-de-sac was counted as an error. Total time per trial was recorded.

The means (and their sigmas) of the conventional total maze scores in learning, including the ten criterial trials, are: For trials, 42 ± 37.5; for errors, 65.4 ± 63.2; for time, 531 ± 368 seconds. The corresponding means for relearning are: For trials, 18.6 ± 8.8; for errors, 6.9 ± 9.7; for time, 108 ± 58.2 seconds. The means for relearning are derived from the records of only twenty subjects; one subject did not return, the other failed to relearn to the criterial ten trials in the time allotted.

The records of the pathways taken in successive trials indicate that the subjects, without exception, tried to follow

the same pathway each time. Of the twenty possible shortest pathways, all but two were traced by a subject at sometime during learning, although only ten of them were finally fixated by the subjects. There is no significant difference between the records of the experienced and inexperienced learners. After the first few trials, the subject traced a perfect pathway and the remaining trials were repeated attempts to follow this same pathway. Errors were made when the stylus went outside the chosen path, the subject often attempting to retrace and find it again. The number of correct trials in the chosen path increased in frequency as the criterion was approached.

In the records of the twenty-two subjects there are 351 errorless runs preceding the criterial ones and *all but 45 (or 13 percent) of these show that the pathway followed was the particular one chosen by the subject and followed by him in the final criterial trials.* Ten subjects took a constant path in all their perfect trials. Eighteen of the subjects made their criterial runs in a constant pathway; of the 40 criterial trials of the remaining four subjects only 9 were in paths other than the one fixated. Only one alternative pathway was taken by each of these four subjects and this was one closely similar to the repeated pathway.

In relearning nineteen subjects made all their correct trials, including the final ten, in the pathway which they had followed repeatedly during learning. The one exceptional subject traced a path differing in two units from his constant one on two trials. The subject who failed to relearn made 39 perfect runs, all in the pathway fixated in learning.

The experimenter recorded verbal reports for each subject, consisting in remarks made during learning and answers to questions put to him after learning, and required each to draw the pattern of the maze at the end of relearning. The following significant facts are revealed by these qualitative data: (1) The drawing was usually a poor reproduction of the one pathway fixated by the subject; (2) the subjects did not feel a tendency to explore the maze after the first few learning trials; (3) the subject was generally up-set and disoriented

when his stylus moved out of the customary pathway; (4) the subject learned under a self-imposed instruction to follow the same path; (5) the subject had very little knowledge of the total maze pattern and only three knew that there was more than one pathway to the goal; (6) the pathway was more or less verbalized by the subject; (7) the subject verbalized the general direction of the goal from the start position.

The fact that there was any variation at all in correct pathways followed by these subjects is evidence that orientation toward the goal was operative in a sense other than merely the final local cue. Its importance, however, varies with the individual learner. In the learning of none of the subjects was orientation to local cues of as little importance as it seems to be in maze learning by rats. The data of this experiment do not offer a basis for inferring which of the two types of orientation is the more important in stylus maze learning. But they do tell us that the one type, orientation to local cues in the pathway, is of relatively greater importance to stylus maze learning than it is to maze learning in rats. Although the stylus maze learning process is influenced by general goal, or directional, sets, the process still fits our concept of a serial habit.

There are three factors which possibly contribute to the making of this difference between the learning of a stylus maze by human subjects and of a similar maze by rodents. In the first place, the habit is largely verbalized in human beings and verbalization is necessarily a more analytical level of behavior than that employed by the rat in a maze. This is a characteristic difference between the two learnings that probably applies in all other habits. Secondly, and perhaps of most importance in the stylus maze situation, the instructional and self-imposed sets determine the relatively great importance of local cues in the pathway for the human subject. Since he knows he must avoid errors in order to make a perfect trial, the subject is oriented toward details of the pathway. This is a factor not operating in hungry rats, for which only the goal (food) is definitely defined by the set. And in the third place, the tendency toward ex-

ploration which is present in rats in a maze does not operate after the first trial or two in human stylus maze learning.

(Manuscript received May 20, 1935)

REFERENCES

1. DASHIELL, J. F., Direction orientation in maze running by the white rat, *Comp. Psychol. Monog.*, 1930, **7**, No. 32.
2. LASHLEY, K. S., Nervous mechanisms in learning, in *A Handbook of General Experimental Psychology*, ed. by C. Murchison, Worcester, Clark Univ. Press, 1934; esp. p. 476.
3. See page 18 in DASHIELL's monograph.

A COMPARISON OF EMPHASIS UPON RIGHT AND UPON WRONG RESPONSES IN LEARNING

BY WILBUR S. HULIN AND DANIEL KATZ

Princeton University

The Problem

Tolman, Hall & Bretnall (**4**) have offered as disproof of the law of effect evidence that an electric shock accompanying a correct response aids learning more than an electric shock accompanying a wrong response. Also they report that the sounding of a bell with a correct response is more effective for learning than the sounding of a bell with a wrong response. Furthermore, the sounding of a bell for right responses is as effective as the bell plus electric shock.

The learning device employed in Tolman, Hall & Bretnall's experiment was the Punch Board Maze. This maze consisted of thirty pairs of holes irregularly placed in a 3′ x 3′ board. One hole of each pair had an electric connection to the shock or bell stimulus. This pattern of electrical connections was used in all experiments except for one check experiment in which the electrical connections were shifted to the alternate hole for each pair.

Goodenough (**1, 3**) criticized Tolman, Hall & Bretnall's results principally on the ground that the differences found could be due to differences in the learning ability of the groups rather than to the methods themselves.

Muenzinger (**2**) repeated and extended the experiment in an attempt to explain the discrepancy between the usual result with animals (in which case shock produces avoidance) and the results with humans found by the above experimenters. In additional series he reversed the pattern of the electrical connections for each of Tolman, Hall & Bretnall's methods; he also added a method of using shock for one hole of a pair and sounding a bell for the other hole of the pair, *e.g.*, a shock for the right response and a bell for the wrong response.

Muenzinger concluded in general that his results did not

show any significant differences between the various experimental conditions, namely, "shock for wrong and shock for right responses, shock for right responses and no shock at all, bell for wrong and bell for right responses, and pattern A [Tolman, Hall & Bretnall's pattern] and pattern B [Muenzinger's reversal of pattern A]."

A comparison of some of Tolman, Hall & Bretnall's and Muenzinger's results follows:

1. In Tolman, Hall & Bretnall's experiment the average error for the Bell-Right group was 13.41 less than for the Bell-Wrong group (*i.e.*, BW 117.22 – BR 103.81). Similarly, in Muenzinger's experiment the average error for the Bell-Right group was 29.1 less than for the Bell-Wrong group (*i.e.*, BW 145.2 – BR 116.1).[1]

2. In Tolman, Hall & Bretnall's experiment the average error for the Bell-Shock-Right group was 23.11 less than for the Bell-Shock-Wrong group (*i.e.*, BSW 128.82 – BSR 105.71). In Muenzinger's experiment the average error for the Bell-Shock-Right group was 25.4 less than for the Bell-Shock-Wrong group (*i.e.*, BSW 143.4 – BSR 118.0).[2]

3. In Tolman, Hall & Bretnall's single comparison of the two patterns (in the Bell-Wrong situation) they found that the average error for pattern B was 8.58 less than the average error for pattern A (*i.e.*, A 117.22 – B 108.64). Muenzinger found (in the Bell-Wrong situation) that the average error for pattern B was 20.6 less than the average error for pattern A (*i.e.*, A 145.2 – B 124.6); however, (in the Bell-Right situation) he found that the average error for pattern A was 9.4 less than for pattern B (*i.e.*, B 125.5 – A 116.1). In two cases pattern B appeared less difficult than A, but in the third case pattern A appeared the less difficult. Thus no consistent difference between the two patterns was shown.

Muenzinger further concludes that the critical ratios[3] show no significant differences between the averages of the

[1] These results were all obtained for pattern A.

[2] Muenzinger's BSW average was obtained for pattern B; the other averages were for pattern A.

[3] The critical ratio is the quotient of the obtained difference divided by its probable error.

groups. The critical ratios for the similar situations in Tolman, Hall & Bretnall's and Muenzinger's experiments follow:

1. The difference in average error for the Bell-Right and the Bell-Wrong groups in Tolman, Hall & Bretnall's experiments was 1.95 times the probable error of the difference. For Muenzinger the difference in average error for the Bell-Right and Bell-Wrong groups was 3.3 times the probable error of the difference.

2. The difference in average error for Tolman, Hall & Bretnall's Bell-Shock-Right and Bell-Shock-Wrong groups was 2.99 times the probable error of the difference. For Muenzinger the difference in average error for the Bell-Shock-Right and Bell-Shock-Wrong [4] was 2.83 [5] times the probable error of the difference.

Although Muenzinger presents critical ratios for other comparisons the present writers regard the above comparisons as those most relevant to the problem.

METHOD

The purpose of the present experiment was to obtain further data upon the learning situations in which the sound of a bell accompanies the correct response (Bell-Right) and in which the sound of a bell accompanies the incorrect response (Bell-Wrong). Since Tolman, Hall & Bretnall as well as Muenzinger found that the bell was as effective as the bell plus electric shock, only the bell was employed in the present study.

The apparatus was a duplicate of Tolman's Punch Board Maze (designed by Robert H. Bruce). The electrical connections were attached to the holes which composed pattern A.

Two groups of 26 students were tested. One group learned the Bell-Right situation; the other group learned the Bell-Wrong situation. Each subject was presented with the same instructions as had been used by Tolman *et al.* Each subject was given three practice trials in which he was instructed to place the stylus in both holes of every pair. After the three practice trials the one group was told to place the stylus in the holes which sounded the bell; the other group was told to place the stylus in the holes which did not sound the bell. The subjects were also instructed to correct an error immediately by placing the stylus in the right hole. The subjects followed the order of pairs of holes as numbered. A testing of the 30 pairs of holes in order constituted one trial. The first trial in which no errors were made was the criterion for learning. Tolman, Hall & Bretnall used the same criterion except for subjects who failed to learn the maze in 20 trials; for these subjects the errors for the 20 trials were counted. In the present experiment only one subject failed to learn the maze within 20 trials; this subject took 22 trials all of which were counted.

[4] Muenzinger used pattern B for BSW.

[5] The value 2.83 was calculated by the present writers from Muenzinger's table.

Results

Table I presents the total number of errors and the total number of trials for each of the 52 subjects under the two

TABLE I

Showing Total Errors and Trials for each S in the Two Groups

Bell-Right			Bell-Wrong		
Subj.	Tot. No. Errors	Tot. No. Trials	Subj.	Tot. No. Errors	Tot. No. Trials
Wa.....	0	1	Si.....	0	1
Ha.....	1	2	Mi.....	19	4
Sm.....	5	4	Fo.....	27	4
Ho.....	17	5	Ba.....	34	13
Rh.....	18	5	En.....	38	10
Ka.....	18	6	Wh.....	47	8
My.....	22	6	Br.....	59	13
Du.....	22	8	Fo.....	64	11
Gi.....	28	7	Hi.....	65	11
Ba.....	31	9	La.....	66	8
Br.....	32	10	Wh.....	67	9
An.....	34	5	Ba.....	67	10
Re.....	40	10	Wi.....	70	13
Co.....	54	11	Re.....	72	12
Sn.....	55	14	Wa.....	76	12
Ba.....	60	12	Ho.....	81	12
Ph.....	66	10	Ma.....	81	17
Wo.....	78	12	Ki.....	87	15
We.....	83	16	St.....	89	11
Dr.....	90	13	Bo.....	90	13
Pu.....	108	14	Co.....	109	17
Fi.....	112	19	Br.....	114	15
Wa.....	118	14	Ta.....	130	15
We.....	119	18	Fr.....	133	17
Ha.....	143	20	Mc.....	143	22
Sl.....	160	18	Ja.....	147	16
Sum.....	1514	269	Sum.....	1975	309
Av.....	58.2	10.3	Av.....	76.0	11.9

experimental situations. For the Bell-Right group the number of trials range from 1 to 20, with an average of 10.3 trials. For the Bell-Wrong group the number of trials range from 1 to 22, with an average of 11.9 trials. In the Bell-Right group the total errors for the individual subjects range from 0 to 160, with an average of 58.2 errors. In the Bell-Wrong group the total errors for the individual subjects range from 0 to 147, with an average of 76.0 errors. The average of the errors for the Bell-Right group is 17.8 less than the average

of the errors for the Bell-Wrong group. This difference is 2.3 times its probable error.

These results show that the Bell-Right situation was more easily learned (according to the criteria of the average number of trials and the average number of errors) than the Bell-Wrong situation. This finding agrees with the following results from Tolman, Hall & Bretnall and from Muenzinger for the average number of errors:

	Average number of errors		Diff.	C.R.
	(Bell-Right)	(Bell-Wrong)		
Tolman, Hall & Bretnall.....	103.81	117.22	13.41	1.95
Muenzinger.....	116.1	145.2	29.1	3.3
Hulin and Katz.....	58.2	76.0	17.8	2.3

The present writers consider that this agreement represents a confirmation of Tolman, Hall & Bretnall's results, and tends to contradict the conclusion proposed by Muenzinger regarding his own data.

(Manuscript received June 5, 1935)

References

1. Goodenough, F. L., A note on Tolman's 'disproof' of Thorndike's law of effect, J. Exper. Psychol., 1933, 16, 459–462.
2. Muenzinger, K. F., Motivation in learning. II. The function of electric shock for right and wrong responses in human subjects, J. Exper. Psychol., 1934, 17, 439–448.
3. Tolman, E. C., The law of effect: a reply to Dr. Goodenough, J. Exper. Psychol., 1933, 16, 463–470.
4. Tolman, E. C., Hall, C. S., and Bretnall, E. P., A disproof of the law of effect and a substitution of the law of emphasis, motivation and disruption, J. Exper. Psychol., 1932, 15, 601–614.